12-19-2007

Implied Measures of Relative Fund Performance

Steve Hogan
Lehman Brothers

Mitch Warachka
Chapman University, warachka@chapman.edu

Recommended Citation

Hogan, S., & Warachka, M. (2008). Implied measures of relative fund performance. *Financial Markets and Portfolio Management, 22*, 47–66. https://doi.org/10.1007/s11408-007-0070-6

Implied Measures of Relative Fund Performance*

Steve Hogan[†] and Mitch Warachka[‡]

September 18, 2006

Abstract

We evaluate the relative performance of funds by conditioning their returns on the cross-section of portfolio characteristics across fund managers. Our implied procedure circumvents the need to specify benchmark returns or peer funds. Instead, fund-specific benchmarks for measuring selection and market timing ability are constructed. This technique is robust to herding as well as window dressing and mitigates survivorship bias. Empirically, the conditional information contained in portfolio weights defined by industry sectors, assets and geographical regions is critically important to the assessment of fund management. For each set of portfolio characteristics, we identify funds with success at either selecting securities or timing-the-market.

*The comments of Bill Ding, Wayne Ferson, Bruce Grundy, Laura Starks and Sheridan Titman along with participants at the European Finance Association 2005 annual meeting significantly improved earlier versions of this paper. All remaining errors are our own. We also thank Morningstar for providing us with the required data. In particular, the efforts of Stephen Murphy at Morningstar are greatly appreciated. Mitch Warachka is grateful for financial support from the SMU-Wharton Research Center, Singapore Management University.

†Lehman Brothers, 745 75^{th} Avenue, New York, N.Y., 10019. Email: stevenchogan@gmail.com

‡Please address correspondence to: Mitch Warachka, #04-01 L.K.C. School of Business, Singapore Management University, 50 Stamford Road, Singapore, 178899.

Email: mitchell@smu.edu.sg Phone: (65) 6828-0249 Fax: (65) 6828-0777

1 Introduction

To circumvent the exogenous specification of benchmarks when evaluating fund managers, we examine their relative performance by constructing fund-specific benchmarks conditioned on *class* portfolio weights. Any characteristics capable of being expressed in terms of portfolio weights can define a class. For example, classes may pertain to investments in distinct industry sectors (healthcare vs. energy), assets (equity vs. bonds) or geographical regions (United States vs. Europe).

Our methodology begins by inferring class-specific returns and variances through a *cross-sectional* regression of fund returns on class-specific portfolio weights whose residuals also depend on these portfolio weights. An individual fund's portfolio weights then convert these implied class-specific returns and variances into a fund-specific benchmark return and variance for evaluating its selection and market timing ability. Selectivity is related to a fund's choice of individual securities within a class, while market timing ability is concerned with a fund manager's portfolio allocation between the classes.[1] Therefore, at every portfolio disclosure date, separate t-statistics for these two attributes are created. An individual fund's performance is then evaluated by examining its *time series* of implied t-statistics. In particular, overperforming fund managers have selectivity and market timing statistics that are consistently positive.

Mamaysky, Spiegel and Zhang (2005) document that performance rankings involving alpha intercepts from a factor model actually reflect the estimation error of their beta coefficients. Relative evaluation avoids this difficulty since classes are not necessarily associated with sources of risk. Instead, our implied returns reflect any risk premiums corresponding to their respective classes.[2] Therefore, any criteria an investor considers relevant to fund selection may define the class portfolio weights. For example, if an investor is seeking to allocate their investments internationally, then the analysis would condition on geographical portfolio weights.[3] A diagnostic test is provided to summarize the importance of conditioning on a particular set of portfolio

[1]For convenience, we refer to funds and fund managers interchangeably although Baks (2003) reports that the former are more important to performance.

[2]Similarly, no-arbitrage pricing focuses on the relative relationship between security prices to avoid imposing assumptions on investor preferences and utility specifications.

[3]Ferson and Schadt (1996) as well as Christopherson, Ferson and Glassman (1998) condition on macroeconomic information when evaluating mutual funds and pension funds respectively.

characteristics.

Our relative evaluation approach is robust to herding and window dressing. Intuitively, herding behavior prevents fund managers from distinguishing themselves from any relative benchmark. Furthermore, transactions between individual securities within a class leave the portfolio weights underlying our analysis unaltered. Therefore, our selectivity measure is invariant to herding and window dressing. By construction, implied returns and variances are free from survivorship bias as they are computed cross-sectionally at a single timepoint. More importantly, provided survival is related to having invested (or not having invested) in certain classes, survivorship bias is mitigated over longer horizons.

Daniel, Grinblatt, Titman and Wermers (1997) list several advantages of using portfolio characteristics to evaluate fund performance, while Chan, Chen and Lakonishok (2002) confirm their superiority over traditional factor sensitivities when predicting returns. The "hypothetical" peer funds in Daniel, Grinblatt, Titman and Wermers (1997) are defined by book-to-market, size and past return quintiles. However, none of these portfolios are traded in the market.[4] In contrast, our methodology is designed for selection decisions between fund managers. Furthermore, conditioning on fund portfolio weights enables us to examine the performance implications of portfolio characteristics with greater precision than classifications derived from quintiles or deciles. Chan, Chen and Lakonishok (2002) report that funds have book-to-market and size properties which are closely aligned with the S&P 500. However, exogenously specifying an appropriate benchmark is a highly contentious issue when evaluating fund performance. For example, the S&P 500 is inappropriate for funds consisting of both equity and bonds (balanced funds), international securities (global or emerging market funds) or those concentrated in specific industries (technology funds). Indeed, the existing literature often invokes benchmarks that imply the majority of fund managers, potentially all of them, can underperform or overperform.

By exploiting the "overlap" in fund investments, the approach of Cohen, Coval and Pástor (2005) motivates our relative evaluation framework. Indeed, commonality in the portfolio weights of different fund managers is the basis for both our procedures. However, we introduce

[4] Appendix A discusses the difference between average returns from characteristic portfolios versus our implied returns in greater detail.

implied performance measures for selection and market timing ability, while Cohen, Coval and Pástor (2005) augment existing performance metrics such as Jensen's alpha. Furthermore, by utilizing class-specific portfolio weights, our methodology is more robust to fluctuations in the holdings of individual assets between disclosure dates.

An empirical illustration of our methodology utilizes a survivorship bias-free set of Morningstar data that consists of portfolio weights for different industry sectors, assets and geographical regions. Of the 1,754 unique funds in our sample, 30% emphasize their *global* or *international* focus, while the majority have non-equity investments. Fund investments are also widely distributed over several industry sectors.[5] Daniel and Titman (1997) posit that the relationship between an asset's expected return and its book-to-market ratio results from firms having similar underlying properties such as common industry or regional exposures. Empirical evidence in Griffin and Karolyi (1998) confirms earlier results by Heston and Rouwenhorst (1994) regarding the distinct roles of industrial and geographical diversification.

Moderate selection and market timing ability is consistently displayed by a small but statistically significant subset of fund managers. However, the "intersection" of moderately overperforming funds across the industry, asset and geographical classifications is nearly empty. Therefore, selection and market timing ability is not diffused across the portfolio characteristics since individual funds rarely overperform across all three criteria. Funds that focus their investments in a small number of classes are more likely to exhibit selectivity at the expense of market timing ability. Overall, fund management skill appears to be specialized. We also document the critical importance of conditioning on portfolio weights when evaluating fund management as our implied performance measure has little in common with its unconditional counterpart that ignores industry, asset and geographical characteristics.

The remainder of this paper begins with the introduction of our evaluation framework in Section 2. Our estimation procedure is then described in Section 3, while properties of the implied performance metrics are discussed in Section 4. Section 5 contains our empirical study

[5]The proposed methodology can refine the performance assessment of fund managers adhering to value (or growth) strategies if classes are defined by book-to-market and industry characteristics. Alternatively, our evaluation procedure could condition on industry portfolio weights but restrict its attention to the cross-section of value (or growth) fund managers.

with Section 6 concluding and offering suggestions for further research.

2 Implied Metrics and Performance Measurement

We begin by highlighting the economics underlying our implied statistics for selection and market timing ability. Classes are designated $c = 1, \ldots, C$ while the funds themselves are indexed by $p = 1, \ldots, P$. Note that we only require class portfolio weights. The holdings of individual securities are unnecessary.

2.1 Simple Regression Interpretation

The intuition behind our relative evaluation procedure is illustrated below with two classes whose portfolio weights are labeled $w_{p,1}$ and $w_{p,2}$. Consider the cross-sectional regression of the aggregate return for fund p on the two portfolio weights

$$r_p \;=\; \beta_1 w_{p,1} + \beta_2 w_{p,2} + \epsilon_p \,, \tag{1}$$

where the expected value of ϵ_p is zero under the null hypothesis of no investment skill. This regression infers $\hat{\beta}_1$ and $\hat{\beta}_2$ which represent implied class returns.[6] These estimates yield a corresponding fund-specific benchmark return

$$\hat{r}_p \;=\; w_{p,1}\,\hat{\beta}_1 + w_{p,2}\,\hat{\beta}_2 \,. \tag{2}$$

This benchmark is customized to the fund's allocation between the two classes through its dependence on $w_{p,1}$ and $w_{p,2}$.

A time series of fund-specific return deviations, $\hat{\epsilon}_p = r_p - \hat{r}_p$, is available since equation (1) yields P residuals, one per fund, at each point in time. Therefore, an individual fund's performance may be assessed by analyzing its $\hat{\epsilon}_p$ time series. As with standard factor models, a fund manager's skill is assessed by analyzing deviations between their observed and "expected" returns. However, instead of calibrating the benchmark expected return using time series data,

[6]Throughout the paper, we adopt the standard convention of denoting parameter estimates with hats. As a result, implied class-specific returns (and variances) are accompanied with hats.

we utilize a cross-sectional procedure which conditions on fund portfolio weights. Overall, the average and standard deviation of the deviations form the ratio

$$\frac{\text{average of } \hat{\epsilon}_p}{\text{standard deviation of } \hat{\epsilon}_p} \, . \tag{3}$$

This simple performance measure is conditioned on fund portfolio weights. Moreover, the relative nature of this technique arises from the cross-sectional estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ underlying the fund-specific benchmark return in equation (2).

The above analysis imposes several restrictions on performance measurement which are overcome by our enhanced methodology. First, there is no distinction between selectivity and market timing ability. Second, the $\hat{\beta}$ coefficients are independent of the class variances. Third, the variance of the $\hat{\epsilon}_p$ residuals is assumed to be constant. However, suppose the first class is cash while the second is equity. Attempts to time-the-market by altering the $w_{p,1}$ and $w_{p,2}$ portfolio weights over time are ignored by the denominator of equation (3).

Before introducing the selectivity and market timing measures, we briefly examine the issue of short-selling implicit in our relative benchmark. Specifically, the $\hat{\beta}_1$ coefficient in equation (2) can be expressed as a linear combination $\{x_1, x_2\}$ of the fund returns

$$\begin{aligned}
\hat{\beta}_1 &= x_1 r_1 + x_2 r_2 && (4) \\
&= x_1 \left(\beta_1 w_{1,1} + \beta_2 w_{1,2} + \epsilon_1 \right) + x_2 \left(\beta_1 w_{2,1} + \beta_2 w_{2,2} + \epsilon_2 \right) \\
&= \beta_1 \left(x_1 w_{1,1} + x_2 w_{2,1} \right) + \beta_2 \left(x_1 w_{1,2} + x_2 w_{2,2} \right) + \left(x_1 \epsilon_1 + x_2 \epsilon_2 \right) \, , && (5)
\end{aligned}$$

where the second equality follows from substituting in equation (1). In order for $\hat{\beta}_1$ to be unbiased, the first and second sums in equation (5) are subject to the following constraints

$$\begin{aligned}
x_1 w_{1,1} + x_2 w_{2,1} &= 1 && (6) \\
x_1 w_{1,2} + x_2 w_{2,2} &= 0 \, , && (7)
\end{aligned}$$

while an unbiased estimate for $\hat{\beta}_2$ requires equations (6) and (7) to equal zero and one respectively. With $w_{p,1}$ and $w_{p,2}$ being positive, a negative value for either x_1 or x_2 is needed to satisfy these constraints. According to equation (4), this property is equivalent to short-selling a fund. However, conditional on knowing the portfolio weights for each fund, this feature of our implied relative benchmark reduces to short-selling individual stocks. Therefore, our relative benchmark return is similar to the HML and SMB factors of Fama and French (1993) or the momentum factor of Carhart (1997) which are constructed using short positions.[7]

2.2 Selectivity Metric

Our first cross-sectional metric evaluates a fund manager's selection ability. We begin our analysis by introducing some additional notation:

- W: a $C \times P$ matrix of observed portfolio weights in class c for fund p.
- $\hat{R}$: a C dimensional implied vector of returns $\hat{r}_c$ for each class.
- $\hat{\Theta}$: a $C \times C$ implied matrix of variances and covariances $\hat{\sigma}_{c,c'}$ for each class.

Conditional on portfolio weights in each class, a cross-sectional procedure described in the next section infers the $\hat{R}$ and $\hat{\Theta}$ parameters. These implied class-specific quantities are the fundamental "building-blocks" of our technique. Therefore, the underlying economic structure of our relative evaluation framework is presented assuming they have been calibrated.

Let $w_{p,c}$ denote the portfolio weight of fund p in class c, with $r_{p,c}$ signifying the *unobservable* return of the fund manager in this class. Although the $r_{p,c}$ returns are not disclosed by funds, their unavailability has no bearing on our analysis since they only serve an intermediary role. More importantly, a fund-specific implied benchmark return

$$\hat{r}_p = \sum_{c=1}^{C} w_{p,c}\,\hat{r}_c \tag{8}$$

is formed by combining a fund's portfolio weights with the implied class-specific returns $\hat{r}_c$ from the $\hat{R}$ vector. The implied returns $\hat{r}_c$ are obtained from a cross-sectional estimation procedure described in the next section, and are critical to the return decompositions which follow below. Although these decompositions may appear trivial at first glance, when combined with our statistical methodology for calibrating $\hat{R}$ and $\hat{\Theta}$, they enable us to evaluate a fund manager against a time-varying fund-specific relative benchmark return.

[7]Book-to-market, size and momentum factors which equally-weight individual stocks essentially underweight or overweight a subset of stocks relative to their value-weighted position in the market portfolio. For example, small growth stocks could be sold-short.

Using the benchmark return in equation (8), the observed aggregate return of a fund is decomposed as[8]

$$
\begin{aligned}
r_p &= \sum_{c=1}^{C} w_{p,c}\, r_{p,c} = \sum_{c=1}^{C} w_{p,c}\, \hat{r}_c + \sum_{c=1}^{C} w_{p,c}\, (r_{p,c} - \hat{r}_c) \\
&= \hat{r}_p + (r_p - \hat{r}_p) \qquad (9) \\
&= \text{implied benchmark without selection ability} + \text{selection ability}\,.
\end{aligned}
$$

Each $r_{p,c} - \hat{r}_c$ deviation results from a fund manager's selection of securities within class c, and ultimately yields the weighted difference $r_p - \hat{r}_p$. In economic terms, a fund manager skilled at selecting individual securities within the various classes has a positive deviation from their implied benchmark. However, the decomposition in equation (9) does not require the unobservable returns $r_{p,c}$ as these terms only serve an intermediary role. The implied benchmark return in equation (8) is written more succinctly as

$$
\hat{r}_p = \mathbf{w}_p^T \hat{R}\,, \qquad (10)
$$

where $\mathbf{w}_p$ is the vector of portfolio weights for fund p in each class. Specifically, $\mathbf{w}_p$ is a column of W containing the portfolio weights $w_{p,c}$ in each of the C classes. Using standard operations from portfolio theory, the corresponding variance of the fund's benchmark return equals

$$
\begin{aligned}
\hat{\sigma}_p^2 &= \sum_{c=1}^{C} w_{p,c}\, \hat{\sigma}_c^2 + \sum_{c=1}^{C} \sum_{c' \neq 1}^{C} w_{p,c}\, w_{p,c'}\, \hat{\sigma}_{c,c'} \\
&= \mathbf{w}_p^T\, \hat{\Theta}\, \mathbf{w}_p\,. \qquad (11)
\end{aligned}
$$

Observe that equations (10) and (11) condition $\hat{r}_p$ and $\hat{\sigma}_p^2$ on a fund's investments in each class since both quantities are functions of $\mathbf{w}_p$. However, $\hat{\sigma}_p$ does not *directly* reflect variability in a

[8]When applying our procedure to funds which engage in short-selling, the classes may distinguish between long and short positions. Although short-selling results in negative portfolio weights, the absolute value of the portfolio weights sum to one. For example, suppose $w_{p,c}$ consists of \$2 and \$1 worth of long and short positions respectively. This portfolio weight is then subdivided into $w_{p,c,L} = 2/3\, w_{p,c}$ and $w_{p,c,S} = -1/3\, w_{p,c}$ with $|w_{p,c,L}| + |w_{p,c,S}| = w_{p,c}$. Distinct implied returns for the long and short positions are then inferred.

fund's return or the returns of individual securities. Instead, $\hat{\sigma}_p$ is related to the cross-section of fund returns. For example, if every fund holds an identical position in each individual security, $\hat{\sigma}_p$ equals zero even when the securities are highly volatile since all fund managers earn the same return. With variability conditioned on class portfolio weights rather than the return fluctuations of an individual fund, equation (11) is appropriate for ascertaining the significance of a fund's deviation from its benchmark return. Therefore, for each fund, the selectivity statistic

$$S_p = \frac{r_p - \hat{r}_p}{\hat{\sigma}_p} \tag{12}$$

is formed. This metric evaluates the deviation between a fund's return and its benchmark, normalized by the benchmark's volatility. Under the null hypothesis of no selection ability, $S_p \overset{d}{\sim} \mathcal{N}(0,1)$ with positive (negative) values indicating overperformance (underperformance).[9]

In comparison to the denominator of equation (3), observe that $\hat{\sigma}_p$ in equation (11) is time-varying. Consequently, a fund may alter its class portfolio weights over time without biasing these cross-sectional metrics. Furthermore, several important differences between the Sharpe ratio and the S_p metric in equation (12) are worth emphasizing. First, the implied benchmark return $\hat{r}_p$ replaces the riskfree rate. Second, $\hat{r}_p$ and $\hat{\sigma}_p$ involve implied parameters. Third, both these quantities are conditioned on a fund's class portfolio weights, hence the return and volatility of the benchmark are fund-specific. Fourth, $\hat{\sigma}_p$ refers to variability in the cross-section of fund returns, not the volatility of individual fund or security returns.

2.3 Market Timing Metric

Besides selecting securities within a class, fund managers also allocate their portfolio across the various classes. In particular, funds that successfully time-the-market earn higher returns by deviating from *benchmark portfolio weights* denoted $w_{B,c}$ which are discussed later in this subsection.

Conditional on fund and benchmark portfolio weights, the benchmark return $\hat{r}_p$ is decom-

[9]More formally, our test statistics have a t-distribution when the number of available funds is insufficient to invoke normality.

posed as

$$\begin{aligned}
\hat{r}_p &= \sum_{c=1}^{C} w_{p,c}\,\hat{r}_c = \sum_{c=1}^{C} w_{B,c}\,\hat{r}_c + \sum_{c=1}^{C} \left(w_{p,c} - w_{B,c}\right)\,\hat{r}_c \\
&= \hat{r}_B + \left(\hat{r}_p - \hat{r}_B\right)\,. \qquad (13)
\end{aligned}$$

The term $\hat{r}_B$ represents the return of a fund manager without market timing or selection ability who simply earns the benchmark return in each class and holds the benchmark portfolio weights. Extending the decomposition in equation (9) using equation (13), a fund's aggregate return is expressed as follows

$$\begin{aligned}
r_p &= \sum_{c=1}^{C} w_{B,c}\,\hat{r}_c + \sum_{c=1}^{C} \left(w_{p,c} - w_{B,c}\right)\,\hat{r}_c + \sum_{c=1}^{C} w_{p,c}\,\left(r_{p,c} - \hat{r}_c\right) \\
&= \hat{r}_B + \left(\hat{r}_p - \hat{r}_B\right) + \left(r_p - \hat{r}_p\right) \qquad (14) \\
&= \text{implied benchmark without market timing or selection ability} \\
&\quad + \text{market timing ability} + \text{selection ability}\,.
\end{aligned}$$

For emphasis, selectivity evaluates the deviation $r_p - \hat{r}_p$ conditional on an individual fund's portfolio characteristics, while market timing considers the difference $\hat{r}_p - \hat{r}_B$ by also conditioning on a set of benchmark portfolio weights. Furthermore, observe that selectivity is independent of the benchmark portfolio weights which define market timing ability. This property ensures a clear distinction between these two attributes.

The benchmark return and variance of the market timing measure, denoted $\hat{r}_B$ and $\hat{\sigma}_B^2$ respectively, equal

$$\begin{aligned}
\hat{r}_B &= \sum_{c=1}^{C} w_{B,c}\,\hat{r}_c \\
&= \mathbf{w}_B^T\,\hat{R}\,, \quad \text{and} \qquad (15)
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}_B^2 &= \sum_{c=1}^{C} \left(w_{p,c} - w_{B,c}\right)\,\hat{\sigma}_c^2 + \sum_{c=1}^{C}\sum_{c'\neq 1}^{C} \left(w_{p,c} - w_{B,c}\right)\,\left(w_{p,c'} - w_{B,c'}\right)\,\hat{\sigma}_{c,c'} \\
&= \left(\mathbf{w}_p - \mathbf{w}_B\right)^T\,\hat{\Theta}\,\left(\mathbf{w}_p - \mathbf{w}_B\right)\,. \qquad (16)
\end{aligned}$$

Equations (15) and (16), when combined with equation (13), yield the following T_p statistic for market timing ability

$$T_p = \frac{\hat{r}_p - \hat{r}_B}{\hat{\sigma}_B}. \tag{17}$$

Under the null hypothesis of no market timing ability, this metric has a $\mathcal{N}(0,1)$ distribution.

For clarification, the variance-covariance matrix $\hat{\Theta}$ is identical for each fund at a particular point in time. By implication, no fund is assumed to possess a more accurate assessment of the covariances between different class returns. This property is consistent with the absence of market timing ability under the null hypothesis since no fund is assumed to have an informational advantage regarding future class performance.[10] Furthermore, when analyzing a fund's selectivity and market timing ability, the return decomposition in equation (14) utilizes the same implied returns. Therefore, a single estimation procedure is sufficient for evaluating both attributes.

Combining portfolio weights with returns to measure market timing ability has also been investigated in Becker, Ferson, Myers and Schill (1999). However, the existing literature examines the correlation between changes in portfolio weights and future returns of individual securities rather than decomposing fund returns as in equation (14). As emphasized in Grinblatt and Titman (1993), this approach is problematic when fund managers exploit return correlation, alter their portfolio's risk across time, or target securities whose expected return and risk have recently risen. These limitations are addressed in Ferson and Khang (2002) who determine a set of benchmark portfolio weights by incorporating publically available information on fund holdings.

The flexibility to specify benchmark portfolio weights is important since $w_{B,c}$ may depend on an investor's objective. As discussed in Chan, Chen and Lakonishok (2002), investors seeking diversification are not well-served by fund managers who attempt to time-the-market by concentrating their investments in a small subset of classes. In the interests of relative evaluation, we consider benchmark portfolio weights defined as

$$w_{B,c} = \frac{1}{P}\sum_{p=1}^{P} w_{p,c} \quad \text{for } c = 1, \ldots, C \tag{18}$$

[10] We thank Wayne Ferson for highlighting to us this property of the null hypothesis as well as motivating the later material on the subject of unconditional versus conditional fund correlation.

which equal the average portfolio weights across funds at a point in time. Regardless of the exact specification for $w_{B,c}$, the return decomposition in equation (14) utilizes implied parameters.

As seen in the next subsection, a time series of selectivity and market timing metrics enables us to measure a fund manager's performance over longer horizons.

2.4 Performance Measurement of Individual Funds

Intuitively, evidence of investment skill requires a fund to consistently exhibit overperformance. For brevity, our current exposition focuses on the selectivity attribute but the statistical technique is immediately applicable to market timing. Let n_p denote the number of observations for fund p during the sample period. As seen from equation (12), each $S_{p,i}$ metric has an i.i.d. $\mathcal{N}(0,1)$ distribution under the null hypothesis of no selection ability.

In practice, a fund manager is often compared to a benchmark index whose volatility is ignored. This situation corresponds to the $K = 0$ threshold since overperformance is indicated by positive $r_p - \hat{r}_p$ deviations, regardless of their variance. Conversely, for $K = 1$, a fund's return is required to exceed its fund-specific benchmark return by one standard deviation in order to exhibit overperformance. More formally, under a binary classification scheme, the number of occurences where $S_{p,i}$ exceeds K is defined as

$$X_p = \sum_{i=1}^{n_p} 1_{\{S_{p,i} > K\}} , \tag{19}$$

implying X_p has a binomial distribution. As a result, our multiperiod test statistic parallels the approach of Agarwal and Naik (2000) when $K = 0$.[11]

Let α represent the probability that $S_{p,i} > K$. Given the distribution of $S_{p,i}$ under the null, the relationship between α and K is available from the standard normal cumulative distribution function as α equals the percentile associated with $\mathcal{N}(0,1) \geq K$. Hence, K and α are used interchangeably to signify the performance threshold.

The associated null hypothesis of no performance is $H_0 : p \leq \alpha$, where p denotes the sample probability that $S_{p,i} > K$, estimated as $\hat{p} = \frac{X_p}{n_p}$. Thus, $\hat{p}$ equals the proportion of

[11]The simplest performance measure would compute the time series average of the implied cross-sectional metrics for an individual fund. However, the time series average is not robust to outliers and cannot ascertain whether an individual fund's $S_{p,i}$ metrics are *consistently* positive (or above a certain performance threshold).

the selectivity statistics that exceed the threshold. The resulting test statistic for selectivity is denoted $B_p(S_{p,i}|\alpha, n_p)$ with a binomial distribution, $Bin(\alpha, n_p)$, under the null. Thus, the corresponding p-value of the performance measure equals the probability that $Bin(\alpha, n_p) \geq X_p$, implying the test statistic is rejected whenever

$$p\text{-value of } B_p(S_{p,i}|\alpha, n_p) \;=\; \sum_{j=X_p}^{n_p} \left(\begin{array}{c} n_p \\ j \end{array} \right) \alpha^j \; (1-\alpha)^{n_p - j} \tag{20}$$

is below its specified Type I error. Equation (20) differentiates between investment skill and luck by requiring a fund manager to consistently exceed their customized benchmark return. In particular, an individual fund's performance is determined by its sequence of implied $S_{p,i}$ metrics.

For large n_p, the following approximation of equation (20) is applicable

$$B_p(S_{p,i}|\alpha, n_p) \;=\; \frac{\hat{p} - \alpha}{\sqrt{\frac{\alpha\,(1-\alpha)}{n_p}}} \;\stackrel{d}{\sim}\; \mathcal{N}(0,1)\,. \tag{21}$$

To examine an individual fund's market timing ability, a $B_p(T_{p,i}|\alpha, n_p)$ test statistic is formed from its time series of $T_{p,i}$ metrics in an identical fashion.[12]

Having introduced the implied selection and market timing metrics as well as their performance measure, the next section details our statistical methodology for inferring class returns and variances.

3 Estimation Procedure

Intuitively, our cross-sectional technique may be expressed in terms of the following "regression"

$$r_p \;=\; \mathbf{w}_p^T\, \hat{R} + \hat{\epsilon}_p \qquad \text{where} \qquad \hat{\epsilon}_p \stackrel{d}{\sim} \mathcal{N}\left(0, \mathbf{w}_p^T\, \hat{\Theta}\, \mathbf{w}_p\right)\,, \tag{22}$$

which is summarized as

$$r_p \;=\; \hat{r}_p + \mathcal{N}\left(0, \hat{\sigma}_p^2\right)\,. \tag{23}$$

[12]To ensure later statistical tests are of the stated significance level when employing the discrete binomial distribution, a randomization correction is incorporated with details in Casella and Berger (1990).

Equation (23) emphasizes the relative nature of our evaluation procedure as only 50% of the funds are able to exceed their benchmark return at a particular point in time. In contrast to equation (1), the $\hat{\epsilon}_p$ residuals in equation (22) are conditioned on class portfolio weights. Furthermore, the $\hat{\sigma}_p^2$ variances are computed cross-sectionally to ensure they are not biased by market timing activity that alters the class portfolio weights over the sample period.

3.1 Likelihood Function

The vector of observed fund returns denoted R_P is assumed to be normally distributed as $\mathcal{N}\left(W^T\hat{R},\, \mathrm{diag}\left\{W^T\hat{\Theta}W\right\}\right)$ where $\mathrm{diag}\left\{W^T\hat{\Theta}W\right\}$ denotes the diagonal of the $W^T\hat{\Theta}W$ matrix, written explicitly as

$$\hat{\Lambda} = \begin{bmatrix} \mathbf{w}_1^T\,\hat{\Theta}\,\mathbf{w}_1 & 0 & \dots & 0 \\ 0 & \mathbf{w}_2^T\,\hat{\Theta}\,\mathbf{w}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{w}_P^T\,\hat{\Theta}\,\mathbf{w}_P \end{bmatrix}.$$

After conditioning on portfolio weights, the residuals $\hat{\epsilon}_p$ in equation (22) are independent but not identically distributed due to their dependence on fund portfolio weights. As emphasized in the next subsection, the conditional independence of the residuals is consistent with considering the diagonal elements of $W^T\hat{\Theta}W$, and found to be appropriate in a later robustness test. Furthermore, this diagonal structure simplifies the maximum likelihood estimation (MLE) of benchmark returns and variances for each class.

Maximizing the conditional log-likelihood of observed fund returns involves maximizing the function

$$\ln L(\hat{R},\hat{\Theta}) \;=\; -\frac{P}{2}\ln(2\,\pi) - \frac{1}{2}\sum_{p=1}^{P}\ln\left(\mathbf{w}_p^T\,\hat{\Theta}\,\mathbf{w}_p\right) - \frac{1}{2}\sum_{p=1}^{P}\frac{\left(r_p - \mathbf{w}_p^T\,\hat{R}\right)^2}{\mathbf{w}_p^T\,\hat{\Theta}\,\mathbf{w}_p}, \tag{24}$$

or simply minimizing

$$\sum_{p=1}^{P}\ln\left(\mathbf{w}_p^T\,\hat{\Theta}\,\mathbf{w}_p\right) + \sum_{p=1}^{P}\frac{\left(r_p - \mathbf{w}_p^T\,\hat{R}\right)^2}{\mathbf{w}_p^T\,\hat{\Theta}\,\mathbf{w}_p} \tag{25}$$

with respect to $\hat{R}$ and $\hat{\Theta}$. Since their solutions are intertwined, an iterative procedure related to the two-stage Fama-MacBeth (1973) regression is necessary.[13] First, given $\hat{\Theta}$, the solution for $\hat{R}$ equals

$$\hat{R} \;=\; \left(W^T \, \hat{\Lambda}^{-1} \, W\right)^{-1} \, W^T \, \hat{\Lambda}^{-1} \, R_P \,, \tag{26}$$

which resembles a weighted least-squares estimator. Second, conditional on $\hat{R}$, elements of the variance-covariance matrix are obtained by minimizing equation (25) with respect to $\hat{\Theta}$. Thus, the following iterative scheme is available:

1. Given $\hat{\Theta}_j$, solve for $\hat{R}_j$ using equation (26).

2. Given $\hat{R}_j$, solve for $\hat{\Theta}_{j+1}$ by minimizing equation (25) using non-linear optimization.

3. Repeat steps 1 and 2 for $j = 0, 1, 2, \ldots$ until convergence is achieved.

The class-specific benchmark parameters $\hat{r}_c$ and $\hat{\sigma}_c^2$ incorporate the returns of any fund with an investment in class c. In other words, every fund which invests in this class contributes to their estimation. By conditioning on class portfolio weights, there is no need to specify peer funds nor are funds required to have their entire portfolio invested in one class. Unlike benchmark returns computed as the average return of individual securities, our implied returns $\hat{r}_c$ are conditioned on actual fund investments. Furthermore, to clarify, the implied return and variance parameters are independent of the benchmark portfolio weights.

3.2 Unconditional versus Conditional Fund Returns

For emphasis, our framework's empirical calibration imposes no structure on the unconditional covariances between observed fund returns, $Cov\,(r_p, r_{p'})$ for $p \neq p'$. Only the *conditional* covariances between fund return deviations (residuals),

$$\hat{\Lambda}_{p,p'} \;=\; Cov\,(r_p - \hat{r}_p,\, r_{p'} - \hat{r}_{p'} \mid \mathbf{w}_p, \mathbf{w}_{p'}) \;, \tag{27}$$

[13] Although estimating $\hat{\Lambda}$ simultaneously from the deviations $r_p - \hat{r}_p$ is circular since $\hat{r}_p$ is conditioned on $\hat{\Lambda}$, the off-diagonal entries of the $\hat{\Lambda}$ matrix may be calibrated from a time series of $\hat{R}$ vectors over the sample period. These elements could then be employed as a weighting matrix. As reinforced by equation (26), this procedure parallels a generalized least squares (GLS) calibration exercise. However, the average pairwise correlation between fund deviations is found to be nearly zero (0.002) in our dataset, while fewer than ten percent of these coefficients are significant at the 5% level according to their bootstrapped confidence intervals and p-values.

are assumed to be zero by the assumption that Λ is diagonal.[14] An identical situation arises in standard linear regression models where the residuals are uncorrelated after having conditioned on independent variables. In our context, the $r_p - \hat{r}_p$ terms are independent while fund returns r_p are allowed to be correlated.

To clarify, the $C \times C$ variance-covariance matrix denoted $\hat{\Theta}$ applies to class returns, while $\hat{\Lambda}$ has a $P \times P$ structure whose off-diagonal elements represent residual correlation between fund returns after having conditioned on their portfolio weights.

3.3 Active versus Passive Fund Management

The value of active fund management may be assessed by our relative evaluation procedure. The Bayesian studies of Pástor and Stambaugh (2002a, 2002b) utilize passive investments to better distinguish fund management skill from model inaccuracy. In our context, two alternatives are available to determine the contribution of active management.

First, passive investments such as exchange-traded funds (ETF's) or index funds may be incorporated into the cross-sectional estimation of $\hat{R}$ and $\hat{\Theta}$. In fact, passive investments could be utilized exclusively when inferring these class-specific returns and variances. A performance comparison between passive versus active management is then facilitated by investigating the subset of overperforming funds. If the overperforming funds are actively managed, then there exist fund managers who possess investment skill.

Second, the implied class returns may be compared with available passive alternatives. For example, if there exists an ETF for class c (such as an industry sector), then comparing this instrument's return with $\hat{r}_c$ would gauge the value of active management.

[14] By definition, the non-random benchmark returns are uncorrelated,

$$Cov\left(\hat{r}_p, \hat{r}_{p'} \mid \mathbf{w}_p, \mathbf{w}_{p'}\right) \quad = \quad Cov\left(E\left[r_p \mid \mathbf{w}_p\right], E\left[r_{p'} \mid \mathbf{w}_{p'}\right] \mid \mathbf{w}_p, \mathbf{w}_{p'}\right) = 0\,.$$

4 Properties of Implied Performance Measures

Relative performance is an appropriate criteria for selecting between fund managers which circumvents the need to specify benchmark returns or peer funds. Cohen, Coval and Pástor (2005) also recognize the importance of exploiting "overlap" in fund investments by augmenting existing performance metrics with the portfolio weights of *individual* securities. However, our class portfolio weights are robust to fluctuations in the holdings of individual securities between portfolio disclosure dates.

Risk premiums associated with the class definitions are captured by our implied estimation procedure, which also avoids the difficulties associated with calibrating time-varying risk premiums. Typically, factor coefficients are calibrated as constants over rolling windows of one to five years. However, Mamaysky, Spiegel and Zhang (2005) document the erroneous fund performance assessments arising from estimation error in the factor coefficients.

Wermers (1999) reports that growth orientated funds often herd. Conflicting empirical evidence on this phenomena is contained in Lakonishok, Shleifer and Vishny (1992) as well as Grinblatt, Titman and Wermers (1995). Fund managers may also engage in window dressing by selling "embarrassing" positions prior to a portfolio disclosure date as documented in Lakonishok, Shleifer, Thaler and Vishny (1991), Musto (1997, 1999) and Carhart, Kaniel, Musto and Reed (2002).

However, the class portfolio weights in our analysis are unaltered by transactions within a class. Thus, our selectivity measure is invariant to herding and window dressing. Indeed, only a small subset of funds could successfully window dress their class portfolio weights without the implied benchmarks being adjusted.[15] Furthermore, selling securities in low return classes and purchasing their high return counterparts compromises a fund's market timing assessment. Specifically, the market timing benchmarks for these funds would condition on high return characteristics despite the relatively low return arising from the fund's actual investments. Thus, window dressing undermines a fund manager's capacity to display market timing ability. Moreover, the tendency for fund managers to misspecify their style after poor performance,

[15]Consequently, even if funds are willing to disclose investments in classes with poor returns, they cannot "reverse" window dress since this behavior would modify the relative benchmarks.

highlighted in Brown and Goetzmann (1997) as well as Chan, Chen and Lakonishok (2002), is alleviated by our approach since fund portfolio weights provide greater objectivity into a fund's investment characteristics.

Although survivorship bias is inherently a database issue, our approach can reduce its impact on performance measurement. As discussed in Brown, Goetzmann, Ibottson and Ross (1992) as well as Carpenter and Lynch (1999), persistence in fund performance may be overstated as a result of heteroskedastic volatility since funds with greater volatility should offer higher returns, provided they survive. However, our methodology is independent of individual fund return volatility since we calibrate cross-sectional return variability.

In contrast, multiperiod persistence tests are biased towards reversals as funds with poor previous performance can only survive by improving. Brown and Goetzmann (1995) as well as Carhart (1997) find empirical evidence consistent with multiperiod survival criteria. However, our selection and market timing ability metrics are inferred from a cross-section of returns and portfolio weights at a single point in time. More importantly, provided survival depends on having invested (or not having invested) in specific classes, our performance measures mitigate survivorship bias since survival is manifested in the time series of S_p and T_p metrics through the portfolio weights. In addition, Brown and Goetzmann (1995) among others find empirical evidence that fund attrition is a consequence of poor returns. Thus, the attrition of poor funds would actually induce a higher implied benchmark return and reduce the likelihood of obtaining overperformance due to survivorship bias.

5 Data and Empirical Results

To illustrate our proposed test procedures, mutual fund data from Morningstar is utilized. Although Morningstar removes defunct funds (as if they never existed) from their published products, our sample is constructed from the original databases and includes all funds with available portfolio weight information.[16] Our sample period is from December 1992 to December 2001. The Morningstar data is ideal for our purposes as fund portfolios are classified according to their weights in specific industries, assets and geographical regions.

[16]We thank Stephen Murphy at Morningstar for constructing the database in this manner.

For industry sectors, Morningstar provides 12 different classes while the original asset categories are reduced to 4 by combining US and non-US equity positions as well as US and non-US bond holdings. Thus, our asset classifications are cash, equity, bonds and a separate class for preferred shares along with convertibles. Morningstar also reports portfolio weights for different geographical regions although very small entries are associated with Central and Latin America, Canada, Africa, Central and Eastern Europe as well as Australia. Therefore, three mergers are enacted to form 5 distinct classes. First, the United States, Canada as well as Central and Latin America are combined into an America class. Second, a Europe class is created by combining Western Europe with Eastern and Central Europe. This class also contains Africa but excludes the United Kingdom which remains a separate entity. Third, Asia is merged with Australia to yield an Asian class which is distinct from Japan.

Since we are interested in evaluating managerial ability, the usual convention of considering gross fund returns without adjustments for fees and expenses is adopted. Performance is computed for each fund with at least 20 observations. For the cross-sectional inference of $\hat{R}$ and $\hat{\Theta}$, a total of 1,601, 1,754 and 1,551 unique funds are available for the respective industry, asset and regional analyses. A summary of the portfolio weights underlying the three classifications is contained in Table 1, which reveals that portfolio weights are evenly distributed across the industry sectors. In contrast, the asset and region portfolio weights are dominated by equity and America respectively.[17]

Additional information is also provided by Morningstar on the expense ratio, size and turnover of each fund. These variables are augmented by a fund's *focus*, defined as the disparity between its portfolio weights

$$\text{Focus}_p = \max_{c=1,\ldots,C} w_{p,c} - \min_{c=1,\ldots,C} w_{p,c}\,, \tag{28}$$

which is an element of the $[0,1]$ interval. Funds which invest equal amounts in each class have zero focus, while those invested exclusively in a single class have a focus measure equaling one.

The time series of class portfolio weight $w_{p,c}$ also has an associated standard deviation $\sigma_{p,c}$ computed over the sample period. The extent to which a fund alters their class portfolio weights

[17] The industry and region portfolio weights only pertain to a fund's equity investments. Adjusting for the amount invested in the remaining three asset classes produces nearly identical results.

is defined as the average

$$\bar{\sigma}_p \quad = \quad \frac{1}{C}\sum_{c=1}^{C}\sigma_{p,c}\,. \tag{29}$$

Thus, $\bar{\sigma}_p$ supplements the fund's reported turnover by only accounting for transactions that alter their class portfolio weights. Indeed, transactions within a class have no influence on equation (29).

To reduce the number of parameters requiring calibration, the off-diagonal entries of $\hat{\Theta}$ are set to zero. In economic terms, this simplification implies that deviations from the benchmark return in one class, say software, are independent of deviations in another class such as financial services.[18] A later robustness test justifies this simplification. Our iterative estimation procedure terminates when the difference in implied parameters between successive iterations is less than 10^{-6}. Starting values for the class-specific returns are the coefficients from a linear regression of the fund's return on its portfolio weights. For the return variances, the regression's standard error serves as the common starting value.

As seen in Table 1, implied returns vary considerably across the various classes. Consequently, the conditional information in these portfolio weights exerts a significant influence on performance evaluation. Interestingly, the regression approach in equation (1) produces less reliable implied class returns whose averages are often extremely small or large. Thus, it is imperative to account for differences between the variances of distinct classes by utilizing our iterative estimation procedure.

Funds in our sample typically have quarterly disclosures causing some variability in the number of funds available each month. Nonetheless, selectivity and market timing statistics are computed monthly. A visual illustration in Figure 1 reveals no discernible differences within the quarterly horizon and confirm our assumption of (conditional) normality.

[18] To clarify, the diagonal assumption imposed on $\hat{\Theta}$ is unrelated to the diagonal nature of $\hat{\Lambda}$. In particular, $\hat{\Theta}$ pertains to the variance-covariance matrix of class-specific implied returns while $\hat{\Lambda}$ refers to deviations in individual fund returns from their fund-specific implied benchmarks.

5.1 Multiple Fund Analysis

The interpretation of our empirical results has to account for multiple Type I errors. For example, if one tests 1,000 funds for investment skill at the 5% level, then 50 false rejections of the null are expected. Therefore, we also test whether the subset of overperforming fund managers is statistically significant.

For a specified threshold, let *Percent* denote the percentage of funds (out of P) with significant test statistics at the γ level. Thus, *Percent* represents the subset of funds with p-values in equation (20) below γ, and is inserted into the following test statistic

$$\frac{Percent - \gamma}{\sqrt{\frac{\gamma\,(1-\gamma)}{P}}} \overset{d}{\sim} \mathcal{N}(0,1)\,. \tag{30}$$

For clarification, the performance measure in equation (20) evaluates an individual fund. When examining multiple funds, equation (30) determines whether the subset of overperforming funds is statistically significant.

Statistically, it is possible for funds to overperform at $K = 1$ but underperform at $K = 0$. For example, a fund's return may be below its benchmark in 75% of the periods but one standard deviation above its benchmark in the remaining periods. This fund's performance is highly variable with frequent underperformance interspersed with dramatic success. To measure consistency in fund performance, funds which overperform when $K = 1$ are required to also overperform at the $K = 0$ threshold.

5.2 Fund Performance

Table 2 reports that selection and market timing ability are detected after conditioning on all three portfolio weight classifications when $K = 0$. Thus, a statistically significant subset of funds exhibit moderate investment skill in both attributes. However, at the higher $K = 1$ threshold, selection ability dissipates, while market timing ability is only detected amongst industry classes.

Correlation between securities within a class may be partially responsible for the decrease in selection ability. Specifically, higher correlation between securities implies greater difficulty in selecting investments that overperform. For example, if securities within the software industry are highly positively correlated, then selectivity is difficult to demonstrate in the software class. One potential explanation for the large decline in market timing ability between the asset and region classes at the higher performance threshold is that fund managers are unwilling (or unable) to dramatically alter their exposure to equity. Indeed, given the dominance of the equity class, Table 2 offers reassurance that our market timing methodology is performing appropriately.[19]

Another interesting feature of Table 2 is that as γ increases from 1% to 10%, the percentage of overperforming funds increases by less than a factor of ten. This insensitivity indicates that investment skill is concentrated in a small number of "exceptional" fund managers. In unreported results available upon request, the magnitude of overperformance for funds exhibiting investment ability ranges from 2.7% to 4.2% per annum on average.

The intersection of significant overperformers across the three characteristics is also studied. According to Table 2, these intersections are nearly empty for both selectivity and market timing. In other words, a fund manager able to successfully select securities (time-the-market) in different industry sectors cannot replicate this skill amongst the asset and regional criteria. We also examine the intersection between selection and market timing ability. Since only moderate investment skill is detected, and overperformance is insensitive to γ, we focus on the $K = 0$ and $\gamma = 0.10$ subsets. Table 3 reports that selection and market timing ability are unrelated. This evidence supports the notion that fund managers specialize in either selecting securities within a sector or allocating their portfolio between various sectors. Later in this section, a Logit analysis offers more insight into the relationship between these attributes in terms of fund characteristics.

The "power" of the test statistic in equation (30) is studied using simulation. This simulation experiment has three objectives. First, correlation between the implied returns of different classes is studied. Second, we examine residual correlation between the $r_p - \hat{r}_p$ deviations of different funds when the $\hat{\epsilon}_p$ errors in equation (22) are not independent. Third, the ability of equation (30) to appropriately account for the Type I errors associated with investigating

[19]A robustness test identifies funds which are most likely to display market timing ability, defined as those with $\bar{\sigma}_p$ values in equation (29) above the industry, asset and regional averages. However, the market timing results for this subset of funds parallel those in Table 2.

multiple funds is analyzed.

Scenarios are examined where the null hypothesis of no overperformance is true as well as false, with further details regarding the simulation procedure in Appendix B. In unreported results available upon request, our procedure is found to be insensitive to correlated class returns or residual correlation among fund returns, while the test statistic accepts and rejects the null hypothesis appropriately.

5.3 Importance of Conditional Information

To examine the importance of a particular set of class portfolio weights to relative fund performance, a naive selection statistic S_p^* defined as

$$S_p^* = \frac{r_p - \bar{r}}{\sigma} \tag{31}$$

is formed, where $\bar{r}$ and σ denote the unconditional mean and standard deviation of fund returns. Unlike the S_p metric in equation (12), S_p^* does not incorporate class portfolio weights or implied returns.[20]

Formally, for a chosen significance level, we define the common performance ratio (abbreviated CPR) as

$$\text{CPR} = \frac{2 \times \text{Funds with both } B_p(S_{p,i}|\alpha, n_p) \text{ and } B_p(S_{p,i}^*|\alpha, n_p) \text{ being significant}}{\text{Funds with significant } B_p(S_{p,i}|\alpha, n_p) + \text{Funds with significant } B_p(S_{p,i}^*|\alpha, n_p)}. \tag{32}$$

The CPR lies within the $[0, 1]$ interval and divides the number of funds in an intersection, which have common conditional and unconditional evaluations, by their sum. If the $S_{p,i}$ and $S_{p,i}^*$ metrics yield identical funds, then the CPR equals one. Conversely, when the conditional and unconditional performance measures have no funds in common, this ratio equals zero. Thus, a lower CPR indicates that the portfolio classification contains more important conditional information. Alternatively, (1 - CPR)% of the funds have their evaluations misspecified when a particular set of portfolio characteristics are ignored.

[20] Investigating the importance of conditional information on market timing ability is difficult as this would involve an allocation between undefined classes. Consequently, we focus our attention on the importance of conditional information to the measurement of selection ability.

We focus our CPR results on the 10% significance level as the overperforming subsets are not sensitive to γ. At the $K = 0$ threshold, few funds have common unconditional and conditional performance measures. In particular, there is only a 27.4% chance that a fund's performance evaluation conditioned on industry characteristics coincides with its unconditional counterpart. The CPR increases slightly for assets to 36.2% and equals 47.2% for regions. Thus, without conditioning on class portfolio weights, the majority of funds would have incorrect performance evaluations.[21]

5.4 Fund Characteristics and Performance

Grinblatt and Titman (1994) find that fund performance is positively related to portfolio turnover but unrelated to size and expenses. Carhart, Carpenter, Lynch and Musto (2002) examine the influence of similar variables on fund performance and report limited evidence of any relationships. We augment this set of variables with Focus_p and $\bar{\sigma}_p$ defined in equations (28) and (29) respectively.

A fund's average expense ratio, size, turnover, focus and fluctuations in its portfolio weights $\bar{\sigma}_p$ in relation to its performance measures are studied using a Logit model

$$y_p = \frac{\exp\left\{\gamma_0 + \gamma_1 \text{ Expense}_p + \gamma_2 \text{ Size}_p + \gamma_3 \text{ Turnover}_p + \gamma_4 \text{ Focus}_p + \gamma_5 \; \bar{\sigma}_p\right\}}{1 + \exp\left\{\gamma_0 + \gamma_1 \text{ Expense}_p + \gamma_2 \text{ Size}_p + \gamma_3 \text{ Turnover}_p + \gamma_4 \text{ Focus}_p + \gamma_5 \; \bar{\sigma}_p\right\}} + \epsilon_p \,, \quad (33)$$

where y_p denotes the p-value of either the selectivity or market timing measure in equation (20) at the $K = 0$ performance threshold. Since the dependent variables are probabilities, and smaller entries coincide with greater overperformance, a positive t-statistic in Table 4 implies an increase in this variable corresponds to a reduction in performance.

Empirically, we find the expense ratio, turnover and size have little influence on fund performance. However, focused funds appear to have greater selectivity within an industry but less ability to time-the-market between different industries. These intuitive conclusions are

[21] The common performance ratios increase for $K = 1$ but remain below 50%. This property is intuitive since funds with higher unconditional overperformance have a greater likelihood of remaining an overperformer after conditioning on the class portfolio weights than those with marginal unconditional overperformance.

supported by the negative and positive t-statistics for selectivity and market timing ability respectively. In addition, focused funds which concentrate their investments in a small number of asset and regional classes are less able to time-the-market which confirms the reasonableness of our market timing formulation.

Besides focus, the other fund characteristic unique to the proposed evaluation framework is $\bar{\sigma}_p$, defined in equation (29) to capture time variation in a fund's class portfolio weights. In agreement with our intuition, its negative t-statistic for market timing ability indicates that funds which alter their industry positions more aggressively have a higher likelihood of exhibiting this attribute.[22]

6 Conclusion

This paper offers a methodology to evaluate the relative ability of fund managers to select securities and time-the-market. Selection and market timing metrics are inferred from a cross-section of fund returns, conditional on portfolio weights in different classes across all fund managers. A separate procedure assesses an individual fund's performance over longer horizons. The class definitions may represent any criteria capable of being expressed in terms of portfolio weights that investors deem relevant to fund selection. A common performance ratio is provided to gauge the importance of conditioning on a specific set of portfolio weights.

Our implied statistics measure relative performance and circumvent the need to specify benchmark returns or peer funds. The resulting performance measure is robust to herding and window dressing by fund managers. Over longer horizons, the effect of survivorship bias on performance measurement is also mitigated, while the calibration of risk premiums associated with time-varying benchmark returns is circumvented.

We investigate classes defined with respect to portfolio weights for different industry sectors, assets and geographical regions. A fund's ability to generate returns that exceed its implied

[22]The lack of significance for $\bar{\sigma}_p$ in explaining market timing ability is explained by its significant negative correlation with focus, -0.441 and -0.678 respectively for assets and regions. When focus is removed from equation (33), the t-statistics for $\bar{\sigma}_p$ become positive and significant (p-values below 10^{-3}). Therefore, as expected, funds with larger $\bar{\sigma}_p$ values have a higher likelihood of timing-the-market which justifies our earlier use of $\bar{\sigma}_p$ in the robustness test of market timing ability.

benchmark and one standard deviation above this threshold are examined. For all three sets of conditional information, empirical evidence finds moderate selection and market timing ability is concentrated in a small number of fund managers. However, funds cannot overperform across all three portfolio classifications, and few funds possess selection as well as market timing ability. Thus, investment skill is specialized. The implied performance metrics appear to be unrelated to a fund's expense ratio, size and turnover. Instead, funds that restrict their investments to a small number of classes and vary their class portfolio weights infrequently are less able to time-the-market, but more likely to exhibit selectivity. Moreover, the conditional information contained in portfolio weights is critically important when evaluating fund performance. Indeed, a failure to condition on industry, asset and regional portfolio characteristics yields highly inaccurate fund evaluations.

Avenues for future research include modifying the proposed statistical methodology for applications to book-to-market, size and past return characteristics. In addition, passive investments such as index funds may also be utilized to assess the value of active management.

References

[1] Agarwal, V., and N. Y. Naik, 2000. Multi-period performance persistence analysis of hedge funds. Journal of Financial and Quantitative Analysis 35, 327-342.

[2] Baks, K. P., 2003. On the performance of mutual fund managers. Working Paper, Emory University.

[3] Becker, C., W. E. Ferson, D. H. Myers, and M. J. Schill, 1999. Conditional market timing with benchmark investors. Journal of Financial Economics 52, 119-148.

[4] Brown, S. J., and W. N. Goetzmann, 1997. Mutual fund styles. Journal of Financial Economics 43, 373-399.

[5] Brown, S. J., and W. N. Goetzmann, 1995. Performance persistence. Journal of Finance 50, 679-698.

[6] Brown, S. J., W. N. Goetzmann, R. G. Ibbotson, and S. A. Ross, 1992. Survivorship bias in performance studies. Review of Financial Studies 5, 553-580.

[7] Carhart, M. M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57-82.

[8] Carhart, M. M., R. Kaniel, D. K. Musto, and A. V. Reed, 2002. Learning for the tape: Evidence of gaming behavior in equity mutual funds. Journal of Finance 52, 1035-1058.

[9] Carhart, M. M., J. N. Carpenter, A. W. Lynch, and D. K. Musto, 2002. Mutual fund survivorship. Review of Financial Studies 15, 1439-1463.

[10] Carpenter, J. N., and A. W. Lynch, 1999. Survivorship bias and attrition effects in measures of performance persistence. Journal of Financial Economics 54, 337-374.

[11] Casella, G., and R. L. Berger, 1990. Statistical Inference. Duxbury Press.

[12] Chan, L. K. C., H. L. Chen, and J. Lakonishok, 2002. On mutual fund investment styles. Review of Financial Studies 15, 1407-1437.

[13] Christopherson, J. A., W. E. Ferson, and D. A. Glassman, 1998. Conditioning manager alphas on economic information: Another look at the persistence of performance. Review of Financial Studies 11, 111-142.

[14] Cohen, R., J. Coval, and L. Pástor, 2005. Judging fund managers by the company they keep. Journal of Finance 60, 1057-1096.

[15] Daniel, K. and S. Titman, 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52, 1-33.

[16] Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997. Measuring mutual fund performance with characteristic-based benchmarks. Journal of Finance 52, 1035-1058.

[17] Fama, E. F., and K. R. French, 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

[18] Fama, E. F., and J. D. MacBeth, 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607-636.

[19] Ferson, W. E., and K. Khang, 2002. Conditional performance measurement using portfolio weights: Evidence for pension funds. Journal of Financial Economics 65, 249-282

[20] Ferson, W. E., and R. W. Schadt, 1996. Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51, 425-461.

[21] Griffin, J. M., and G. A. Karolyi, 1998. Another look at the role of industrial structure of markets for international diversification strategies. Journal of Financial Economics 50, 351-373.

[22] Grinblatt, M., and S. Titman, 1993. Performance measurement without benchmarks: An examination of mutual fund returns. Journal of Business 66, 47-68.

[23] Grinblatt, M., and S. Titman, 1994. A study of monthly mutual fund returns and performance evaluation techniques. Journal of Financial and Quantitative Analysis 29, 419-444.

[24] Grinblatt, M., S. Titman, and R. Wermers, 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. American Economic Review 85, 1088-1105.

[25] Heston, S. L., and K. G. Rouwenhorst, 1994. Does industrial structure explain the benefits of international diversification? Journal of Financial Economics 36, 3-27.

[26] Lakonishok, J., A. Shleifer, R. Thaler, and R. W. Vishny, 1991. Window dressing by pension fund managers. American Economic Review 81, 227-231.

[27] Lakonishok, J., A. Shleifer, and R. W. Vishny, 1992. The impact of institutional trading on stock prices. Journal of Financial Economics 32, 23-44.

[28] Musto, D. K., 1997. Portfolio disclosures and year-end price shifts. Journal of Finance 52, 1563-1588.

[29] Musto, D. K., 1999. Investment decisions depend on portfolio disclosures. Journal of Finance 54, 935-952.

[30] Pástor, L., and R. F. Stambaugh, 2002a. Mutual fund performance and seemingly unrelated assets. Journal of Financial Economics 63, 315-349.

[31] Pástor, L., and R. F. Stambaugh, 2002b. Investing in equity mutual funds. Journal of Financial Economics 63, 351-380.

[32] Speigel, M. I., H. Mamaysky, and H. Zhang, 2005. Improved forecasting of mutual fund alphas and betas. Working Paper, Yale University.

[33] Wermers, R., 1999. Mutual fund herding and the impact on stock prices. Journal of Finance 54, 581-622.

Appendices

A Utilizing Information on Individual Securities

Suppose class c consists of N_c individual securities indexed by $k = 1, 2, \ldots, N_c$ whose returns are denoted r_k. This appendix highlights the difference between class-specific expected returns and variances computed as

$$\bar{r}_c = \frac{1}{N_c} \sum_{k=1}^{N_c} r_k \quad \text{and} \tag{34}$$

$$\sigma_c^2 = \frac{1}{N_c - 1} \sum_{k=1}^{N_c} \left(r_k - \bar{r}_c \right)^2 , \tag{35}$$

versus $\hat{r}_c$ and $\hat{\sigma}_c^2$ obtained from our implied statistical procedure. In particular, the above estimates fail to measure relative performance for three reasons. First, $\bar{r}_c$ and σ_c^2 in equations (34) and (35) are not conditioned on the portfolio weights of actual funds. For example, individual securities that are not held by any fund can nonetheless influence these estimates. Second, by implication, the majority of funds could overperform (underperform) any benchmark constructed from the $\bar{r}_c$ estimates in equation (34). Third, the variance in equation (35) is not between fund returns but pertains to the variability of individual securities. For example, σ_c^2 is positive even if all fund managers maintain identical portfolios and produce identical returns as a consequence.

Nonetheless, given market values for the individual securities, class portfolio weights can be constructed from the portfolio weights of individual securities after sorting them according to a specified criteria such as industry SIC codes. Once this initial step is completed, our implied estimation procedure is applicable.

B Simulation Study

The simulation analysis has $C = 12$ classes with a matrix W of portfolio weights sampled from the data for one thousand funds, $P = 1,000$. Random class-specific returns denoted R are drawn from a $\mathcal{N}(R_C, \Theta_C)$ distribution, which creates the vector of fund returns $R_P = W^T R$.

Under the null hypothesis, scenarios in which the matrix Θ_C is both diagonal and dense are evaluated. In the latter case, correlated class returns are instilled into the analysis. Correlation between individual fund returns, which exists even after conditioning on portfolio weights, is also introduced by augmenting the R_p vector as follows

$$R_P^* = R_P + \epsilon^* \quad \text{where} \quad \epsilon^* \sim \mathcal{N}(\mathbf{0}, \Lambda_P) .$$

Thus, R_P^* represents fund returns generated by portfolio weights as well as a random component determined by the variance-covariance matrix denoted Λ_P. Along with W and R_C, the diagonal elements of Θ_C as well as Λ_P originate from the industry data and are fixed throughout the simulation study. Three separate structures for Θ_C and Λ_P are investigated. In the first instance, both of these variance-covariance matrices are diagonal. In the second and third scenarios, Θ_C and Λ_P are dense matrices respectively with a full complement of off-diagonal (covariance) terms. However, in all three cases, variances along the diagonal are identical.

A time series of $n_p = 24$ return vectors are simulated. The estimation procedure then produces $\hat{R}$ and $\hat{\Theta}$ estimates under the previous assumptions that $\hat{\Theta}$ and $\hat{\Lambda} = \text{diag}\left\{W^T \hat{\Theta} W\right\}$ are both diagonal as detailed in Section 3. These estimates are then converted into selectivity metrics before applying the performance measure in equation (20) to all 1,000 funds, yielding a *Percent* subset. The simulation process is repeated $N = 1,000$ times. Finally, each of the N subsets is then tested according to equation (30) for γ equal to 0.05 and 0.10. Besides simulating under the null hypothesis of no selection ability, 15% of funds have their returns increased by a factor of 1.25. In this case, the null hypothesis of no selection ability is false at the $K = 0$ threshold and its rejection is anticipated.

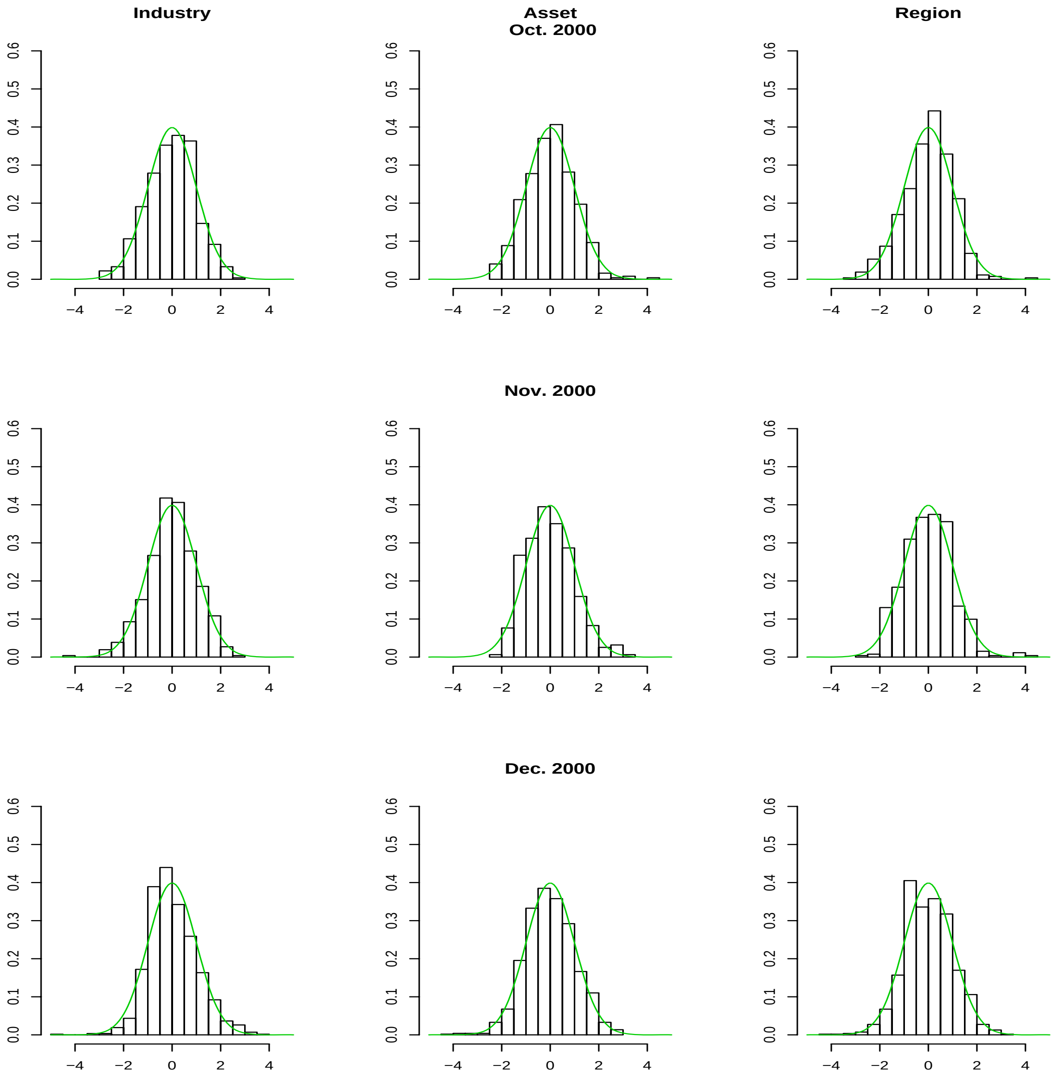

Figure 1: Implied distributions of selectivity statistics for three consecutive months during the sample period. The exact dates correspond to October, November and December of 2000. A normal distribution (with the same mean and variance) is superimposed on each histogram for ease of comparison.

Table 1: Summary statistics for Morningstar data on industries, assets and regions. The mean and median of the class portfolio weights across all funds are reported, along with their standard deviation representing variability across funds. Note that the average portfolio weights sum to one. The annualized mean and median of the implied fund returns for each class are also presented over the December 1992 to December 2001 sample period.

		Class Portfolio Weights w_c			Implied Returns $\hat{r}_c$	
Industries	12 Classes	Mean	Median	Std. Dev.	Mean	Median
	1. Software	4.1	2.9	1.7	30.22	20.30
	2. Hardware	13.5	11.3	4.9	30.94	26.20
	3. Media	4.0	3.0	1.6	24.21	22.32
	4. Telecommunications	6.2	4.7	2.3	39.71	31.32
	5. Healthcare	10.7	10.0	2.8	16.52	17.87
	6. Consumer Services	8.0	7.1	2.5	14.18	13.30
	7. Business Services	6.4	4.8	2.1	16.68	10.93
	8. Financial Services	17.2	15.5	3.5	9.49	3.57
	9. Consumer Goods	9.1	8.4	2.7	14.77	15.94
	10. Industrial Materials	11.7	10.1	2.8	9.80	11.52
	11. Energy	5.6	5.0	2.0	8.87	9.88
	12. Utilities	3.5	1.5	1.4	14.52	14.65
Assets	4 Classes					
	1. Cash	5.4	3.6	4.2	5.70	3.75
	2. Equity	87.8	94.2	5.0	20.23	14.99
	3. Bonds	4.7	6.0	1.7	5.58	5.13
	4. Preferreds and Convertibles	2.1	1.8	1.7	15.08	14.60
Regions	5 Classes					
	1. America	82.3	97.3	1.4	15.70	19.66
	2. United Kingdom	3.1	3.4	0.7	7.84	8.35
	3. Europe	7.7	1.4	1.2	4.07	8.15
	4. Japan	3.3	3.5	0.6	1.07	3.26
	5. Asia	3.6	4.2	0.6	3.26	9.68

Table 2: Summary of selection and market timing performance with classes defined in terms of industries, assets and regions. For $K = 0$, a fund's ability to exceed its fund-specific implied benchmark return is examined, while $K = 1$ ascertains performance one standard deviation above this threshold. The percentages reported below document the proportion of funds that exhibit investment skill by having p-values for the performance measure in equation (20) below the stated significance levels. The asterices *, ** and *** indicate statistical significance of these subsets at the 10%, 5% and 1% level respectively according to equation (30). The row entitled "All Criteria" corresponds to an intersection over all three sets of conditional information, and examines the existence of funds capable of overperforming across each portfolio characteristic.

Percentage of Overperforming Funds

$K = 0$ Performance Threshold	Selectivity			Market Timing		
	Significance Level			Significance Level		
Conditional Information	10%	5%	1%	10%	5%	1%
Industry	13.0***	11.2***	8.6***	11.5*	9.4***	5.5***
Asset	11.5*	10.3***	7.2***	36.7***	29.0***	14.9***
Region	7.3	6.2	4.6***	9.1	7.7***	7.3***
All Criteria	0.5	0.5	0.0	0.7	0.5	0.0

$K = 1$ Performance Threshold	Selectivity			Market Timing		
	Significance Level			Significance Level		
Conditional Information	10%	5%	1%	10%	5%	1%
Industry	2.7	1.9	1.3	4.1	3.2	2.0***
Asset	3.2	2.2	1.5*	0.0	0.0	0.0
Region	2.0	1.3	0.4	0.9	0.9	0.9
All Criteria	0.2	0.2	0.0	0.0	0.0	0.0

Table 3: Success of funds in selecting securities and timing-the-market conditional on industry, asset and regional portfolio weights. The results below examine the intersection of individual funds with significant performance measures for selection and market timing ability at the 10% level. This analysis determines whether a fund's success in selecting securities within the classes is duplicated by their allocation decisions between classes. The reported percentages below correspond to the $K = 0$ performance threshold which evaluates a fund manager's ability to exceed their fund-specific benchmark return. None of these entries are significant at the 10% significance level.

Intersection of Selection and Market Timing Ability Subsets			
		Selectivity	
	Industry	Asset	Region
Market Timing			
Industry	2.1	3.9	1.6
Asset	5.9	7.7	2.9
Region	2.2	1.6	0.7

Table 4: Relationships between selection and market timing ability versus fund characteristics. The results displayed below are reported for the Logit model in equation (33). Focus and $\bar{\sigma}_p$ are defined in equations (28) and (29) respectively. The p-values are reported in parentheses below the t-statistics with *, ** and *** denoting significance at the 10%, 5% and 1% level respectively. Since the dependent variables are the p-values of the performance metrics, smaller p-values imply more significant overperformance. Therefore, a positive (negative) t-statistic implies that larger variables coincide with a decreased (increased) likelihood of overperformance.

Selection Ability

	Intercept	Expense	Size	Turnover	Focus	$\bar{\sigma}_p$
Industry	2.14**	-1.91*	1.39	1.27	-3.47***	0.35
(p-value)	(0.032)	(0.056)	(0.166)	(0.205)	(0.000)	(0.725)
Asset	-1.40	-1.19	0.83	0.39	1.16	3.96***
(p-value)	(0.163)	(0.235)	(0.401)	(0.695)	(0.245)	(0.003)
Region	-0.48	-0.39	1.76*	0.90	0.24	0.21
(p-value)	(0.632)	(0.695)	(0.080)	(0.371)	(0.811)	(0.834)

Market Timing Ability

	Intercept	Expense	Size	Turnover	Focus	$\bar{\sigma}_p$
Industry	2.39***	-0.98	0.38	-0.15	3.79***	-10.08***
(p-value)	(0.017)	(0.328)	(0.700)	(0.883)	(0.000)	(0.000)
Asset	116.15***	0.31	0.74	1.06	-40.20***	-1.57
(p-value)	(0.000)	(0.755)	(0.460)	(0.289)	(0.000)	(0.116)
Region	47.40***	-1.76	0.23	0.58	-6.37***	-0.34
(p-value)	(0.000)	(0.173)	(0.818)	(0.395)	(0.000)	(0.735)