

DISCUSSION PAPER SERIES

No. 3717

JUDGING FUND MANAGERS BY THE COMPANY THEY KEEP

Randolph Cohen, Joshua Coval
and Lubos Pástor

FINANCIAL ECONOMICS



Centre for Economic Policy Research

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP3717.asp

JUDGING FUND MANAGERS BY THE COMPANY THEY KEEP

Randolph Cohen, Harvard Business School
Joshua Coval, Harvard Business School
Lubos Pástor, University of Chicago and CEPR

Discussion Paper No. 3717
January 2003

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Randolph Cohen, Joshua Coval and Lubos Pástor

CEPR Discussion Paper No. 3717

January 2003

ABSTRACT

Judging Fund Managers by the Company They Keep*

We develop a performance evaluation approach in which a fund manager's skill is judged by the extent to which his investment decisions resemble the decisions of managers with distinguished performance records. The proposed performance measures are estimated more precisely than standard measures, because they use historical returns and holdings of many funds to evaluate the performance of a single fund. According to one of our measures, funds with significantly positive ability considerably outnumber funds with significantly negative ability at the end of our sample. Simulations demonstrate that our measures are particularly useful in ranking managers. In an application that relies on such ranking, we find only weak persistence in the performance of US equity funds after accounting for momentum in stock returns.

JEL Classification: G10

Keywords: mutual funds, performance evaluation and persistence

Randolph Cohen
Harvard Business School
Soldiers Field
Boston, MA 02163
USA
Tel: (1 617) 495 6674
Email: rcohen@hbs.edu

Joshua Coval
Harvard Business School
Soldiers Field
Boston, MA 02163
USA
Email: jcoval@hbs.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=135768

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=135768

Lubos Pástor
University of Chicago
Graduate School of Business
1101 East 58th St
Chicago IL 60637
USA
Tel: (1 773) 834 4080
Fax: (1 773) 702 0458
Email: lubos.pastor@gsb.uchicago.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=147096

*Research support from the Center for Research in Security Prices and the James S Kemper Faculty Research Fund at the Graduate School of Business, University of Chicago is gratefully acknowledged (Pástor). We thank George Chacko, Aida Charoenrook, Gene Fama, Owen Lamont, Rob Stambaugh, Pietro Veronesi, Tuomo Vuolteenaho, and seminar participants at Arizona State University, Dartmouth College, Emory University, Harvard University, Indiana University, Michigan State University, Northwestern University, University of Chicago, University of Pennsylvania, and Vanderbilt University for helpful comments. Huong Trieu provided excellent research assistance.

Submitted 03 December 2002

Introduction

People are often judged by the company they keep. Various human characteristics that are difficult to observe directly are commonly inferred from the characteristics of others who behave in a similar manner. For example, a bored air traveler eager to chat about Madonna's life story is more likely to turn to a fellow passenger reading the National Enquirer than to a passenger reading the Wall Street Journal, because the readership of the former periodical is generally known to be more interested in celebrity gossip.

When people compete, their success is often predicted by their techniques, and the quality of a given technique is frequently evaluated based on the track records of the technique's followers. As an example, suppose a group of basketball players, some of whom shoot with both hands and some with only one hand, have been taking 10 shots each at the basket. So far, the average score of the two-handers is $8/10$, while the one-handers' average is only $4/10$. Two players, one one-hander and one two-hander, have completed only half of their shots so far, and both have scored $4/5$. Suppose you are to bet on which of the two players is going to achieve a higher score out of 10. Although the track records of both players are identical, it seems sensible to bet on the two-hander, because the track records of the other players show that two-handed shooters tend to score higher. The one-hander, who is employing what appears to be an inferior technique, is more likely to have been lucky in his first five shots.

Similar to basketball players, active mutual fund managers rely on a variety of techniques when trying to beat their benchmarks, and many of these techniques are common to groups of managers. For example, managers collect information from different sources and use different valuation methods, but there are clusters of managers who use similar sources and similar methods. Managers using similar techniques are likely to make similar investment decisions and also to have similar performance. This paper focuses on the aspect of performance that is due to commonality in the managers' techniques.

In this paper, a fund manager's ability to select outperforming stocks is judged by the extent to which his investment decisions resemble those of other successful managers. One way to assess the similarity of the managers' investment decisions is to compare the compositions of their portfolios. For example, consider two managers with equally impressive past returns, where one manager currently keeps a big chunk of his portfolio

in the stock of Intel, while the other manager holds mostly Microsoft. Suppose also that Intel is currently held especially by managers with good past performance, whereas Microsoft is held mostly by managers with undistinguished records. It seems reasonable to rule that the first manager, whose decision to hold Intel is shared by a higher-caliber set of managers, has superior ability to select stocks, while the second manager, whose techniques coincide with those of subpar managers, has been merely fortunate.

The performance measure proposed in the paper is defined with respect to some simpler reference measure, for which we choose the traditional Jensen's alpha. We show that our measure of a manager's skill is a weighted average of the traditional skill measures across all managers, where the weights are essentially the covariances between the manager's portfolio weights and the weights of the other managers. Put differently, if two managers have highly similar portfolio weights, then one manager's skill contributes substantially to our measure of the other manager's skill.

Another way of comparing the managers' investment decisions is to compare their trades. A modified "trade-based" version of our performance measure judges a manager's skill by the extent to which recent changes in his holdings match those of managers with outstanding past performance. This measure is also a weighted average of the traditional skill measures, but now the weights are essentially the covariances between the concurrent changes in the manager's portfolio weights and those of the other managers. According to the trade-based measure, the manager is skilled if he tends to buy stocks that are concurrently purchased by other managers who have performed well, and sell stocks that are concurrently purchased by managers who have performed poorly.

The extent to which fund managers tend to buy and sell stocks at the same time is analyzed by Lakonishok, Shleifer, and Vishny (1992), Grinblatt, Titman, and Wermers (1995), and Wermers (1999), among others. However, this literature on "herding" does not propose using the degree of similarity in the managers' trading activities to evaluate their performance. Hong, Kubik, and Stein (2002) show that fund managers from the same city are more likely to hold, buy, or sell the same stock at the same time than are managers from different cities. Given this evidence, managers from the same city are likely to be assigned more similar skill according to our measure than managers from different cities. This seems appropriate. Hong, Kubik, and Stein attribute their finding to word-of-mouth information diffusion, and it seems sensible to assign similar skill to

managers using similar information.

Evaluating mutual fund performance is a topic of enormous relevance for the well-being of individual investors and for market efficiency in general. The traditional approach to performance evaluation relies solely on historical fund returns to construct measures such as Jensen's alpha (Jensen, 1968) or Sharpe ratio (Sharpe, 1966). Since return histories of many existing funds are short, these traditional measures are often imprecise. To cope with the low precision, recent studies propose alternative performance measures that rely also on fund holdings.¹ Those measures employ the portfolio holdings of the fund whose performance is being evaluated, but they do not exploit the information about the fund's performance contained in the holdings and past returns of other funds. Including that additional information and documenting its benefits is a contribution of this paper.

Our performance measures offer substantially higher precision compared to the traditional return-based measures. Since our measures are weighted averages of the traditional measures, precision is added by pooling across funds the information contained in the funds' average returns. That is, instead of using just the historical returns of a given manager to estimate his performance, our measures use the return series of all managers whose holdings (or changes in holdings) overlap with those of the given manager. The dramatic improvement in precision is confirmed empirically using the Spectrum mutual fund database. Among the U.S. equity funds at the end of our sample, our holdings-based measure is about four to eight times more precise than the traditional measure, on average, and our trade-based measure is about seven times more precise. The biggest precision gains are obtained for short-history funds, as expected. Overall, our measures are more precise for 90 to 98 percent of all funds.

We find some interesting results regarding the significance of managerial stock-picking skill. According to our holdings-based measure, funds with significantly positive ability before costs and fees vastly outnumber funds with significantly negative ability at the end of our sample. Significantly positive ability is found for 14 to 359 times as many funds as significantly negative ability, depending on the choice of the reference measure.

¹See, for example, Grinblatt and Titman (1993), Daniel, Grinblatt, Titman, and Wermers (1997), Wermers (2000), and Ferson and Khang (2002). With no reliance on holdings data, Pástor and Stambaugh (2002) and Busse and Irvine (2002) show that additional precision in the traditional performance measures can be achieved by incorporating returns on seemingly unrelated passive assets.

Under the four benchmarks of Carhart (1997), only three funds exhibit significantly negative ability, but 1,076 funds (one out of three) exhibit significantly positive ability. This result seems more plausible than inferences based on standard measures as well as on our trade-based measure, which both imply approximately equal proportions of managers with positive and negative skill. While some managers might possess the ability to successfully trade on private information, managers with negative true skill (i.e. managers less skilled than monkeys) must surely be far more difficult to find. A similar argument is made by Baks, Metrick, and Wachter (2001), who proclaim a prior belief that managerial skill before costs and fees cannot be negative.

A simulation analysis is conducted to examine the extent to which our skill measures are able to capture true skill. When managers are ranked by true skill and then separately by various performance estimators, our estimators produce higher rank correlations with true skill than standard estimators that do not exploit similarities in holdings across managers. Our measures are particularly effective when the number of managers is large and when their portfolios differ substantially from each other, i.e. when there is more information to pool across funds. Our measures are shown to exhibit some bias in estimating the traditional measure of skill, Jensen's alpha. Due to the weight-averaging of skill across managers, our holdings-based measure is biased towards the average level of skill in the population of managers, similar in spirit to a Bayesian shrinkage estimator. The nature of the bias does not impair the rank-ordering of managers, though, and our estimators dominate the usual estimators of alpha even when the objective is to rank managers by their alphas. In sum, the simulations reveal that our performance measures are especially useful in applications that involve ranking managers.

One such application is provided in our empirical examination of persistence in the performance of U.S. equity funds. The literature is ambiguous as to whether past mutual fund returns predict future returns. Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995), Elton, Gruber, and Blake (1996), and Bollen and Busse (2002), among others, all find some persistence in mutual fund performance, especially among the funds lagging their benchmarks. However, much of the persistence has been attributed to momentum in stock returns and to fund expenses (Carhart, 1997, and Wermers, 1997), and some also to survivorship bias (Brown, Goetzmann, Ibbotson, and Ross, 1992). The mixed nature of the evidence (see also Malkiel, 1995) begs for further research.

To analyze performance persistence, at the end of each quarter between the years 1977 and 2000, funds are sorted into deciles according to our measures as well as some traditional measures of performance estimated over the past 12 months. The deciles' returns are tracked over the subsequent quarter, and the risk-adjusted performance of the decile portfolios is compared over the full sample. When the traditional estimates of the CAPM and Fama-French alphas are used to rank managers, fund returns exhibit significant persistence: The difference between the risk-adjusted returns of the top and bottom deciles is 8.3% per year for the CAPM alphas and 7.2% for the Fama-French alphas, with the t -statistics of 3.0 and 3.5, respectively. However, the return difference drops to an insignificant 1.4% when the momentum benchmark is included, confirming the results of Carhart (1997) and Wermers (1997). Persistence is slightly stronger when measured by our holdings-based measure: The top-bottom decile return difference is significant at 8.6% per year using the CAPM alpha and 8.3% using the Fama-French alpha as a reference measure. After accounting for momentum, however, the difference again dips below significance, to 1.6%. Our trade-based measure produces smaller return differences with higher t -statistics, but even this measure leads to marginally insignificant persistence after accounting for momentum. The evidence in favor of persistence further weakens when funds are sorted by past performance estimated over 24 months or longer.

We then extend our investigation of persistence by analyzing whether our performance measures are able to predict their own future values. Every year between 1976 and 2000, funds are sorted according to various measures of performance estimated over the previous year, and persistence is assessed by computing the correlations between the fund rankings in adjacent years. When funds are sorted on their usual CAPM alphas, these correlations are significantly positive on average, indicating significant persistence in performance, but the correlations turn insignificant after adjusting for the Fama-French factors and momentum in stock returns. The image of weak persistence is reinforced when funds are sorted on our performance measures. Using the CAPM alpha as the reference measure, our holdings-based measure as well as our trade-based measure exhibit significant persistence, but the significance evaporates when the Fama-French and four-factor alphas are used as reference measures. Persistence in skill is even weaker (though still positive) when measured at a two-year frequency. In short, persistence in fund performance seems hard to find, even with our performance measures, which seem better suited for this investigation than standard measures.

The paper proceeds as follows. Section I introduces our performance measures. Section II discusses a simulation exercise that evaluates the usefulness of these measures in capturing true skill. Section III implements the measures empirically to assess their precision and significance, as well as to investigate persistence in the performance of U.S. equity funds. Section IV concludes.

I New Performance Measures

This section introduces a performance measure that judges a fund manager’s ability by the extent to which his stock holdings overlap with those of managers whose other investments have been successful. After discussing the interpretation and the precision of this measure, the section proposes its modified version, which focuses on changes in fund holdings rather than their levels.

I.1 A Measure Based on Levels of Holdings

Assume there are M managers, $m = 1, \dots, M$, and N stocks, $n = 1, \dots, N$, each of which is held by at least one manager. Let $w_{m,n}$ denote the current weight on stock n in manager m ’s portfolio, and let δ_m denote the skill of manager m . For each stock n , define its quality measure $\bar{\delta}_n$ as

$$\bar{\delta}_n = \sum_{m=1}^M v_{m,n} \delta_m, \tag{1}$$

where

$$v_{m,n} = \frac{w_{m,n}}{\sum_{m=1}^M w_{m,n}}. \tag{2}$$

The quality of stock n is defined as the average skill of all managers who hold stock n in their portfolios, weighted by how much of the stock they hold. Stocks with high quality are those that are held mostly by highly-skilled managers. Managers who hold stocks of high quality are likely to be skilled, because their investment decisions are similar to those of other skilled managers (i.e. such managers are in “good company”). Since a larger position in a stock of given quality reveals more about the manager’s ability, the

population version of our performance measure is constructed as

$$\delta_m^* = \sum_{n=1}^N w_{m,n} \bar{\delta}_n. \quad (3)$$

In words, our measure of a manager's performance is the average quality of all stocks in the manager's portfolio, where each stock contributes according to its portfolio weight.

The population measure δ_m^* relies on true skill δ_m . Since true skill is unknown, our measure must be defined in relation to some reference measure of skill. The reference measure needs to reflect the fund's average return in excess of the return on the appropriate benchmark, which captures any style effects for which the manager should not be rewarded. One sensible choice would be the characteristic-based benchmark of Daniel, Grinblatt, Titman, and Wermers (1997), for example. In the interest of simplicity, we opt for a more conventional reference measure, the traditional Jensen's alpha, defined as the intercept from the regression of the fund's excess returns on benchmark returns. To construct our estimator of managerial skill, we replace δ_m in equation (1) by $\hat{\delta}_m$, the usual OLS estimator of alpha:

$$\hat{\delta}_m^* = \sum_{n=1}^N w_{m,n} \bar{\delta}_n, \quad (4)$$

where

$$\bar{\delta}_n = \sum_{m=1}^M v_{m,n} \hat{\delta}_m. \quad (5)$$

This estimator can in principle be further iterated, as $\hat{\delta}_m^*$ can itself be used in place of $\hat{\delta}_m$ in equation (5). In simulations as well as in the empirical work, we examine not only the estimator in equation (4), but also its iterated counterpart.² Also note that in the unlikely event of no overlap in holdings, when manager m holds only stocks that are held by no other manager, our measure collapses to the traditional measure, $\hat{\delta}_m^* = \hat{\delta}_m$.

Our performance measure has an interesting alternative interpretation. To obtain it, first let h_n denote the ratio of the dollar value of stock n held by all M managers to the total dollar value of all stocks held by these managers. Then

$$h_n = \frac{\sum_{m=1}^M w_{m,n}}{\sum_{n=1}^N \sum_{m=1}^M w_{m,n}} = \frac{\sum_{m=1}^M w_{m,n}}{M}, \quad (6)$$

²Note that the gains from repeated iteration are in principle limited, because replacing δ_m in equation (1) by δ_m^* results in an unsatisfactory definition of δ_m^* . Such a circular definition can be shown to imply that the δ_m^* 's for all managers are equal to the same arbitrary constant. Hence, we iterate only once.

a relation that will be useful shortly. Let W denote the $N \times M$ matrix whose (n, m) element is $w_{m,n}$, let V denote the $N \times M$ matrix whose (n, m) element is $v_{m,n}$, let $\hat{\delta}$ denote the $M \times 1$ vector of $\{\hat{\delta}_m\}_{m=1}^M$, let $\bar{\delta}$ denote the $N \times 1$ vector of $\{\bar{\delta}_n\}_{n=1}^N$, and let $\hat{\delta}^*$ denote the $M \times 1$ vector of $\{\hat{\delta}_m^*\}_{m=1}^M$. Since $\hat{\delta}^* = W'\bar{\delta}$ and $\bar{\delta} = V\hat{\delta}$, the vector of our performance measures from equation (4) can be written as

$$\hat{\delta}^* = Z\hat{\delta}, \quad (7)$$

where $Z = W'V$. The performance of manager m is thus

$$\hat{\delta}_m^* = \sum_{j=1}^M z_{m,j} \hat{\delta}_j, \quad (8)$$

where $z_{m,j}$ is the (m, j) element of Z . Using equations (2) and (6), we have

$$z_{m,j} = \sum_{n=1}^N w_{m,n} v_{j,n} = \frac{1}{M} \sum_{n=1}^N w_{m,n} w_{j,n} \frac{1}{h_n}. \quad (9)$$

Our measure of manager m 's skill, $\hat{\delta}_m^*$, is therefore a weighted average of the usual skill measures across all managers.³ The weight assigned to the performance of manager j , $z_{m,j}$, is a loose measure of covariance between the weights of managers m and j . This is sensible – if managers m and j own many of the same stocks, they may be using similar strategies, and we want to pay a lot of attention to the performance of manager j when evaluating manager m . The scaling factor $1/h_n$ in equation (9) downweights stocks that receive large weights on average in the portfolios of all managers. For example, if a certain stock occupied 20% of the market capitalization, many managers would have large weights in that stock, and its contribution to the performance measures would be overstated in the absence of the scaling factor. Similarly, if both managers happen to hold a lot of a stock that few others hold, that is valuable information and the scaling factor emphasizes it.

Due to the symmetry of Z ($z_{i,j} = z_{j,i}$), which follows from equation (9), it is easy to show that the averages of $\hat{\delta}_m^*$ and $\hat{\delta}_m$ across managers are equal:

$$\frac{1}{M} \sum_{i=1}^M \hat{\delta}_i^* = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^M z_{i,j} \hat{\delta}_j = \frac{1}{M} \sum_{j=1}^M \hat{\delta}_j \sum_{i=1}^M z_{i,j} = \frac{1}{M} \sum_{j=1}^M \hat{\delta}_j. \quad (10)$$

³The weights sum to one:

$$\sum_{j=1}^M z_{m,j} = \sum_{j=1}^M \sum_{n=1}^N w_{m,n} w_{j,n} \frac{1}{M h_n} = \sum_{n=1}^N \frac{1}{M h_n} w_{m,n} \sum_{j=1}^M w_{j,n} = \sum_{n=1}^N w_{m,n} = 1.$$

As a result, our skill measure provides only as much information as the usual measure about the performance of the mutual fund industry as a whole. Nevertheless, our measure can be useful in evaluating the funds' relative performance, as documented below.

The precision of our estimators can be assessed by their squared standard errors, which are displayed along the main diagonal of the $M \times M$ covariance matrix of $\hat{\delta}^*$. This matrix can be computed from equation (7) as

$$\text{Cov}(\hat{\delta}^*, \hat{\delta}^{*'}) = Z\Omega Z', \quad (11)$$

where $\Omega = \text{Cov}(\hat{\delta}, \hat{\delta}')$ is the $M \times M$ covariance matrix of $\hat{\delta}$. If all funds had return histories spanning the same time period, Ω could in principle be calculated from one big multivariate regression of fund excess returns on benchmark returns. However, funds have histories of unequal lengths spanning different periods. The appendix describes how Ω can be calculated in such an environment.

To assess the gains in precision provided by our measure, write the squared standard error of our estimator for manager m as

$$\text{Var}(\hat{\delta}_m^*) = z_m \Omega z_m' = \sum_{i=1}^M \sum_{j=1}^M z_{m,i} z_{m,j} \omega_{i,j}, \quad (12)$$

where z_m is the m -th row of Z and $\omega_{i,j}$ is the (i, j) element of Ω . For simplicity, assume (with some loss of generality) that all m elements of $\hat{\delta}$ have the same standard error, so that $\omega_{i,i} = \omega^2$ for all $i = 1, \dots, M$. Then $\omega_{i,j} = \omega^2 \rho_{i,j}$, where $\rho_{i,j}$ is the correlation between $\hat{\delta}_i$ and $\hat{\delta}_j$. Also assume that all $z_{m,i} > 0$, which is likely to hold in practice for most though not all pairs of funds. Then

$$\text{Var}(\hat{\delta}_m^*) = \sum_{i=1}^M z_{m,i}^2 \omega^2 + \sum_{i=1}^M \sum_{j=1, j \neq i}^M z_{m,i} z_{m,j} \omega^2 \rho_{i,j} \quad (13)$$

$$\leq \omega^2 \left(\sum_{i=1}^M z_{m,i}^2 + \sum_{i=1}^M \sum_{j=1, j \neq i}^M z_{m,i} z_{m,j} \right) = \omega^2 \left(\sum_{i=1}^M z_{m,i} \right)^2 = \omega^2, \quad (14)$$

because we already showed that the rows of Z sum to one. This means that so long as the $\hat{\delta}$'s are not perfectly correlated, $\hat{\delta}_m^*$ has a lower standard error than $\hat{\delta}_m$.⁴ Our gains

⁴If all elements of $\hat{\delta}$ do not have the same standard error, this relation holds only on average and there are likely to be some m 's for which it does not hold. The calculation in equation (14) is analogous to computing the variance of a portfolio of stocks (with no short positions), which is typically smaller than the variance of any given stock in the portfolio.

in precision relative to $\hat{\delta}$ therefore come from the imperfect correlations between the $\hat{\delta}$'s. These correlations tend to be low when the cross-sectional correlation of the managers' residual returns is low. Thus, increasing the number of benchmarks that define $\hat{\delta}$ can improve the precision of our estimator not only by increasing the precision of $\hat{\delta}$, but also by reducing the residual correlations among funds. This observation is confirmed empirically in Section III.

I.2 A Measure Based on Changes in Holdings

In the previous subsection, managers are inferred to be making similar investment decisions if they have similar holdings, regardless of the timing of their trades. In this subsection, we assume that managers make similar decisions if their trades are similar. Since trading involves transaction costs, decisions to trade are likely to reflect stronger views than decisions to passively hold. The performance measure developed here exploits similarities between changes in the managers' holdings, rather than their levels.

The return on the portfolio of manager m at time t can be written as

$$R_{m,t} = \sum_{n=1}^N w_{m,n} r_{n,t}, \quad (15)$$

where $r_{n,t}$ denotes the return on stock n . Define the change in the weights as

$$d_{m,n} = w_{m,n,t} - w_{m,n,t-1} \frac{1 + r_{n,t}}{1 + R_{m,t}}, \quad (16)$$

which is the difference between the current weight and the weight obtained if the manager neither bought nor sold any of this stock over the past period.⁵ Let $\mathcal{N}_m^+ = \{n : d_{m,n} > 0\}$ denote the set of stocks purchased by manager m between $t - 1$ and t and $\mathcal{N}_m^- = \{n : d_{m,n} < 0\}$ denote the set of stocks sold by manager m over the same time period. Analogously, let $\mathcal{M}_n^+ = \{m : d_{m,n} > 0\}$ denote the set of managers who made net purchases of stock n between $t - 1$ and t , and let $\mathcal{M}_n^- = \{m : d_{m,n} < 0\}$ denote the set of managers with net sales. We normalize the changes in the weights as

$$x_{m,n}^+ = \frac{d_{m,n}}{\sum_{n \in \mathcal{N}_m^+} d_{m,n}}, \quad x_{m,n}^- = \frac{d_{m,n}}{\sum_{n \in \mathcal{N}_m^-} d_{m,n}}, \quad (17)$$

⁵In our empirical analysis, one period equals one quarter. If the manager did not trade at all over the past quarter, so that $d_{m,n} = 0$ for all n , our measure is undefined for this manager. Also note that the time subscripts on d , w , and some related measures below are suppressed to simplify notation.

$$y_{m,n}^+ = \frac{d_{m,n}}{\sum_{m \in \mathcal{M}_n^+} d_{m,n}}, \quad y_{m,n}^- = \frac{d_{m,n}}{\sum_{m \in \mathcal{M}_n^-} d_{m,n}}, \quad (18)$$

so that $x_{m,n}^+$ ($x_{m,n}^-$) captures the fraction of manager m 's purchases (sales) accounted for by stock n , and $y_{m,n}^+$ ($y_{m,n}^-$) captures the fraction of purchases (sales) of stock n accounted for by manager m .

For each stock n , define its quality measure $\bar{\delta}_n$ as

$$\bar{\delta}_n = \bar{\delta}_n^+ - \bar{\delta}_n^-, \quad (19)$$

where

$$\bar{\delta}_n^+ = \sum_{m \in \mathcal{M}_n^+} y_{m,n}^+ \hat{\delta}_m \quad (20)$$

$$\bar{\delta}_n^- = \sum_{m \in \mathcal{M}_n^-} y_{m,n}^- \hat{\delta}_m, \quad (21)$$

and $\hat{\delta}_m$ is the usual performance measure, acting as a reference measure again. (Using δ_m in place of $\hat{\delta}_m$ yields the population version of our modified skill measure, δ_m^{**} .) The quality of stock n is the difference between the average skill of all managers who bought stock n recently ($\bar{\delta}_n^+$) and the average skill of all managers who sold stock n recently ($\bar{\delta}_n^-$), where the averages are weighted by how much was bought or sold. Stocks of high quality are those that were recently bought mostly by high-skill managers and sold mostly by low-skill managers. Managers who recently bought high-quality stocks and sold low-quality stocks are likely to be skilled, because their techniques seem similar to those of other skilled managers. Hence, our modified skill measure is

$$\hat{\delta}_m^{**} = \hat{\delta}_m^+ - \hat{\delta}_m^-, \quad (22)$$

where

$$\hat{\delta}_m^+ = \sum_{n \in \mathcal{N}_m^+} x_{m,n}^+ \bar{\delta}_n \quad (23)$$

$$\hat{\delta}_m^- = \sum_{n \in \mathcal{N}_m^-} x_{m,n}^- \bar{\delta}_n. \quad (24)$$

This is the difference between the average quality of the stocks recently bought by manager m and the average quality of the stocks recently sold by this manager. Note that $\hat{\delta}_m^+$ captures to what extent the manager has been buying high-quality stocks and $\hat{\delta}_m^-$

captures to what extent the manager has been selling low-quality stocks. Our measure combines these two aspects of stock-picking skill.

To get more insight into the new measure, we write it out as

$$\begin{aligned}\hat{\delta}_m^{**} &= \sum_{n \in \mathcal{N}_m^+} x_{m,n}^+ \left(\sum_{j \in \mathcal{M}_n^+} y_{j,n}^+ \hat{\delta}_j - \sum_{j \in \mathcal{M}_n^-} y_{j,n}^- \hat{\delta}_j \right) - \sum_{n \in \mathcal{N}_m^-} x_{m,n}^- \left(\sum_{j \in \mathcal{M}_n^+} y_{j,n}^+ \hat{\delta}_j - \sum_{j \in \mathcal{M}_n^-} y_{j,n}^- \hat{\delta}_j \right) \\ &= \sum_{j=1}^M c_{m,j} \hat{\delta}_j,\end{aligned}\tag{25}$$

where

$$\begin{aligned}c_{m,j} &= \sum_{n=1}^N \left[x_{m,n}^+ y_{j,n}^+ \mathbf{1}_{\{n \in \mathcal{N}_m^+\}} \mathbf{1}_{\{j \in \mathcal{M}_n^+\}} - x_{m,n}^+ y_{j,n}^- \mathbf{1}_{\{n \in \mathcal{N}_m^+\}} \mathbf{1}_{\{j \in \mathcal{M}_n^-\}} \cdots \right. \\ &\quad \left. - x_{m,n}^- y_{j,n}^+ \mathbf{1}_{\{n \in \mathcal{N}_m^-\}} \mathbf{1}_{\{j \in \mathcal{M}_n^+\}} + x_{m,n}^- y_{j,n}^- \mathbf{1}_{\{n \in \mathcal{N}_m^-\}} \mathbf{1}_{\{j \in \mathcal{M}_n^-\}} \right],\end{aligned}\tag{26}$$

and $\mathbf{1}_{\Omega}$ denotes an indicator function equal to one or zero depending on whether the associated condition is true. The new measure is therefore a “weighted average” of the usual measures across all managers. The quotation marks are appropriate because the weights sum to zero, $\sum_{j=1}^M c_{m,j} = 0$, as shown in the appendix. The weight on manager j essentially reflects the covariance of the weight changes of managers m and j . This weight, $c_{m,j}$, is a sum of N terms, one for each stock, which capture the products of the managers’ weight changes in that stock. If both managers traded that stock in the same direction (i.e. if both bought or sold it), this product is positive and increasing in the extent of the joint action. If one manager bought and the other sold, the product is negative. Hence the loose covariance interpretation. The higher the covariance of the weight changes of managers m and j , the more attention we want to pay to the skill of manager j when evaluating the skill of manager m .⁶

To summarize the modified measure, let C denote the $M \times M$ matrix whose (i, j) element is $c_{i,j}$. Also, let X^+ and X^- denote $M \times N$ matrices whose (i, j) elements are $x_{i,j}^+$ and $x_{i,j}^-$, respectively. Similarly, let Y^+ and Y^- denote $M \times N$ matrices whose (i, j) elements are $y_{i,j}^+$ and $y_{i,j}^-$, respectively. Zeros are substituted for all (i, j) elements of X^+ and Y^+ such that $d_{i,j} < 0$, as well as for all (i, j) elements of X^- and Y^- such that $d_{i,j} > 0$. With this notation, C can be expressed as

$$C = X^+(Y^+)' - X^+(Y^-)' - X^-(Y^+)' + X^-(Y^-)'.\tag{27}$$

⁶The average of $\hat{\delta}_m^{**}$ across managers can be shown to equal zero in the special case when managers trade stocks only with each other. Since managers generally trade also with other entities such as individuals, the average $\hat{\delta}_m^{**}$ is likely to be close to zero but not exactly zero.

Then the $M \times 1$ vector of our performance measures in equation (22) can be written as

$$\hat{\delta}^{**} = C\hat{\delta}, \quad (28)$$

and the precision of the changes measure can be calculated as

$$\text{Cov}(\hat{\delta}^{**}, \hat{\delta}^{**'}) = C\Omega C', \quad (29)$$

where $\Omega = \text{Cov}(\hat{\delta}, \hat{\delta}')$ as before. Henceforth, we often refer to the measure $\hat{\delta}^*$ as the “levels” measure and to the measure $\hat{\delta}^{**}$ as the “changes” measure.

Our changes measure in equation (22) weighs stock qualities by the relative magnitudes of the weight changes across stocks. There is an interesting alternative definition, $\hat{\delta}_m^{**A} = \sum_{n=1}^N d_{m,n} \bar{\delta}_n$, which instead relies on the absolute magnitudes of these changes. We have explored both changes measures, and found that they lead to the same conclusions throughout the paper. (When $\hat{\delta}^{**}$ is replaced by $\hat{\delta}_m^{**A}$, the results are very similar in all tables except for Table 3, in which $\hat{\delta}_m^{**A}$ has even higher precision than $\hat{\delta}^{**}$). We chose to focus on $\hat{\delta}_m^{**}$ mostly due to its appealing interpretation as the difference between the average qualities of stocks bought and sold.

II Simulations

The purpose of the simulation analysis is to assess the usefulness of our estimators in capturing true managerial skill, which is conveniently known in our simulated environment. When managers are ranked by their true skill and separately by various performance estimators, useful information about estimator quality is provided by the resulting rank order correlations. As shown below, our estimators contain important information about true skill that is not captured by other commonly used metrics. We also rank managers by the traditional estimator of their skill, δ_m , and investigate the extent to which the resulting ranking is similar to the rankings produced by our measures δ_m^* and δ_m^{**} and their estimators. Again, our measures are shown to dominate the standard estimators that make no use of similarities in holdings across managers.

II.1 Simulation Design

The simulations consider a simple setting in which M managers receive signals about expected excess returns of N stocks. Each stock's excess return can be written as

$$r_n = \alpha_n + e_n, \quad n = 1, \dots, N \quad (30)$$

where α_n is stock n 's expected excess return and e_n is an error term. We simulate α_n from $N(0, \sigma_\alpha^2)$ and e_n from $N(0, \sigma_e^2)$. Both α_n and e_n are assumed to be uncorrelated across stocks, which simplifies the calculations and substantially speeds up the simulations.

Each manager receives a signal about each stock. With probability γ_m , this signal is equal to the stock's true alpha, and otherwise it is equal to a noise term drawn from an identical distribution. That is, manager m 's signal about stock n is

$$s_{m,n} = \begin{cases} \alpha_n & \text{with probability } \gamma_m \\ u_n & \text{with probability } 1 - \gamma_m \end{cases} \quad (31)$$

where $u_n \sim N(0, \sigma_\alpha^2)$. Since higher γ_m means higher signal quality, γ_m captures the manager's true skill in this simulated environment. Each manager's γ_m is drawn as

$$\gamma_m = q\bar{g} + (1 - q)g_m, \quad m = 1, \dots, M \quad (32)$$

where \bar{g} and g_m are drawn from the standard uniform distribution. Higher values of q imply that the managers' gammas are more alike, which ultimately leads to more similar portfolio holdings. Lower values of q make the managers more different from each other in terms of their skill. We show below that more heterogeneity across managers increases the benefits from using our performance estimator.

The managers are assumed to know the signal structure in equation (31) as well as their own skill γ_m and the error volatility σ_e . They have no information about α_n other than the signal, and they learn about α_n from the signal using the Bayes rule. As shown in the appendix, the expected return on stock n perceived by manager m is

$$E_{m,n} = \gamma_m s_{m,n}, \quad (33)$$

the perceived variance of stock n 's returns is

$$V_{m,n} = \sigma_e^2 + \sigma_\alpha^2 + \gamma_m(s_{m,n}^2 - \sigma_\alpha^2) - s_{m,n}^2 \gamma_m^2, \quad (34)$$

and the perceived return covariance matrix is diagonal. The managers are assumed to maximize the Sharpe ratios of their portfolios, so that manager m 's weight in stock n is

$$w_{m,n} \propto V_{m,n}^{-1} E_{m,n}. \quad (35)$$

As mutual funds are typically not allowed to short, we also require no short sales. This constraint can be imposed simply by putting a zero weight on any stock with a negative signal, thanks to our assumption of uncorrelated stock returns.

The traditional measure of the true performance of manager m is his expected abnormal return:

$$\delta_m = w'_m \alpha, \quad (36)$$

where α is the $N \times 1$ vector of stock alphas.⁷ Next, we calculate six different performance estimators. Since returns are simulated in one period only, the traditional estimator of δ , $\hat{\delta}$, is calculated simply as the manager's realized return:

$$\hat{\delta}_m = w'_m r, \quad (37)$$

where r is the $N \times 1$ vector of realized excess stock returns r_n . Our performance measure based on the levels of holdings, $\hat{\delta}^*$, is calculated as

$$\hat{\delta}_m^* = Z_m \hat{\delta}, \quad (38)$$

where Z_m is defined as row m of matrix Z in equation (7).⁸ Our performance measure based on the changes in holdings, $\hat{\delta}^{**}$, is calculated as

$$\hat{\delta}_m^{**} = C_m \hat{\delta}, \quad (39)$$

where C_m is defined as row m of matrix C in equation (28). Since the simulation is set in a one-period world, changes in holdings are calculated assuming that each investor initially holds an equally-weighted portfolio of all stocks.

Equations (38) and (39) use $\hat{\delta}$ as the simple initial estimator (or reference measure) that forms an input to our estimator. In principle, $\hat{\delta}$ can be replaced by any sensible

⁷For simplicity, we use no benchmark in the simulations, so that returns coincide with abnormal returns. Simulating benchmark returns seems like an unnecessary complication that would unlikely make much difference, other than to make our assumption of uncorrelated e_n more realistic.

⁸If there happens to be a stock that is not held by any manager, the corresponding row of the $N \times M$ weight matrix W is deleted in calculating $\hat{\delta}^*$, to prevent division by zero in equation (2).

performance estimator, including ours. Therefore, we also compute iterated versions of our estimators, in which $\hat{\delta}$ in equations (38) and (39) is replaced by $\hat{\delta}^*$ and $\hat{\delta}^{**}$:

$$\hat{\delta}_m^{*i} = Z_m \hat{\delta}^* = Z_m Z \hat{\delta} \quad (40)$$

and

$$\hat{\delta}_m^{**i} = C_m \hat{\delta}^{**} = C_m C \hat{\delta}. \quad (41)$$

For comparison purposes, we also construct a simple Bayesian estimator that shrinks $\hat{\delta}$ towards a common mean:

$$\hat{\delta}_m^B = \frac{1}{2} \hat{\delta}_m + \frac{1}{2} \bar{\hat{\delta}}, \quad (42)$$

where $\bar{\hat{\delta}}$ is the average of $\hat{\delta}$'s across all managers.

In addition to the above six estimators, which rely only on holdings and return information, we also calculate the population versions of δ^* and δ^{**} for each manager:

$$\delta_m^* = Z_m \delta \quad (43)$$

$$\delta_m^{**} = C_m \delta, \quad (44)$$

where δ is the $N \times 1$ vector of the true δ_m 's. Since these measures reflect information that is generally unknown outside the world of simulations, they possess an unfair advantage over the six estimators. These measures are computed to assess the maximum achievable gains from using our estimators.

We conduct 10,000 simulations for each set of parameter values. The number of managers M takes on the values of 10, 50, 100, and 300, the number of stocks $N = 10, 50,$ and 100, and the weight $q = 0$ and 0.50. Throughout, $\sigma_\alpha = 0.1$ and $\sigma_e = 0.2$. In each simulated sample, we calculate $\delta, \hat{\delta}, \hat{\delta}^B, \hat{\delta}^*, \hat{\delta}^{*i}, \hat{\delta}^{**}, \hat{\delta}^{**i}, \delta^*,$ and δ^{**} for each manager. The managers are ranked according to each of these measures as well as according to their true skill γ for the purpose of computing the rank order correlations.

II.2 Simulation Results

Table 1 judges the ability of various performance measures to imitate the ranking of managers by their true skill γ . The table reports the rank order correlations of each

estimator with γ , averaged across 10,000 simulations. Importantly, our levels estimator $\hat{\delta}^*$ consistently delivers a higher rank correlation with γ than the traditional estimator $\hat{\delta}$. The only exceptions occur in some cases when the number of managers is very small ($M = 10$), i.e. when there is little cross-sectional information to pool. Our measure is particularly effective when the number of managers is large and when their abilities (and hence also their portfolios) vary significantly. For example, with 300 managers, 50 stocks, and zero weight on the common signal ($q = 0$), our measure delivers a rank correlation of 0.87 whereas $\hat{\delta}$ delivers only 0.67. Iterating generally further improves the levels estimator – as we go from $\hat{\delta}^*$ to $\hat{\delta}^{*i}$, the rank correlations increase unless the number of managers is small. The changes estimator $\hat{\delta}^{**}$ produces even slightly higher rank correlations with true ability than the levels estimator $\hat{\delta}^*$. The population versions of all three measures, δ , δ^* and δ^{**} , produce similar rank correlations, but our measures (especially δ^{**}) tend to prevail when the number of managers is large. Finally, note that as the number of stocks N increases, the managers’ portfolios become better diversified and it becomes easier for all measures to detect skill. Overall, our performance measures appear to be very effective at ranking managers according to their true skill.

Table 2 evaluates the capacity of various skill measures to capture the traditional skill measure δ . The average rank correlations of each estimator with δ , reported in Panel A, lead to the same conclusions as Table 1. Our levels estimator $\hat{\delta}^*$ achieves a higher rank correlation with δ than the traditional estimator $\hat{\delta}$ and the Bayesian estimator $\hat{\delta}^B$ (both of which deliver the same rank correlations by construction), despite the fact that, unlike $\hat{\delta}^*$, those two estimators are designed to capture δ . The only exceptions occur when the number of managers is small, $M = 10$, as before. The effectiveness of our measure again increases with M and decreases with q . With $M = 300$, $N = 50$, and $q = 0$, our measure’s rank correlation with δ is 0.95, but it is only 0.75 for the other two measures. The performance of the changes estimator $\hat{\delta}^{**}$ is similar to that of the levels estimator and, as before, iterating typically helps both estimators unless M is small. The population versions of our measures, δ^* and δ^{**} , which present an upper bound on how close we can hope to get to δ with our sample estimators, attain very high rank correlations (sometimes as high as 0.99) under all specifications. This suggests that the bias in our measures with respect to δ has little effect on manager rankings. To sum up, our estimators seem more successful than the standard estimators in imitating the rankings produced by the true value of the traditional performance measure.

Panel B of Table 2 reports the mean squared errors (MSEs) in estimating δ for all performance measures.⁹ The MSEs are calculated as averages across managers and simulations of the squared differences between the estimate of δ and the true value. As expected, the MSE of the Bayesian estimator $\hat{\delta}^B$ is invariably lower than the MSE of the traditional estimator $\hat{\delta}$. More interestingly, our levels estimator $\hat{\delta}^*$ has a lower MSE than the traditional estimator in all cases except $N = 100$ and $q = 0$, despite the fact that it is not designed to capture δ but δ^* . While $\hat{\delta}^*$ has a lower MSE than $\hat{\delta}^B$ when $N = 10$, the Bayesian estimator generally beats our estimator for larger N . The MSE for the changes estimator $\hat{\delta}^{**}$ is sometimes large, which is not unexpected, since this estimator fails to capture δ even on average, as explained below. Iterating tends to worsen the MSE for both of our estimators. The MSEs of our population measures δ^* and δ^{**} reveal the extent to which these measures depart from the true delta. In general, while our estimators can boast high precision, they also exhibit some bias with respect to δ , so the overall effect on the MSE is unclear.

Figure 1 plots the bias in estimating δ for four performance estimators, $\hat{\delta}$, $\hat{\delta}^B$, $\hat{\delta}^*$, and $\hat{\delta}^{**}$. In each simulated sample, managers are ranked in ascending order by their δ . Let δ_m denote the true performance of the m -th ranked manager, and $\tilde{\delta}_m$ denote that manager's estimated performance (i.e. $\tilde{\delta}_m = \hat{\delta}_m, \hat{\delta}_m^B, \hat{\delta}_m^*$, or $\hat{\delta}_m^{**}$). For each rank m ($m = 1, \dots, M$), the bias is computed by averaging $\tilde{\delta}_m - \delta_m$ across 10,000 simulated samples. The figure plots the bias against the rank m .

The figure's four panels, corresponding to different values of M and q , all tell the same story. The traditional estimator $\hat{\delta}$ is unbiased, whereas both the Bayesian estimator $\hat{\delta}^B$ and our levels estimator $\hat{\delta}^*$ are biased. For the managers with above-average skill, both estimators are biased downwards, and for the managers with below-average skill, both are biased upwards. This bias is entirely expected as a result of the shrinkage of the individual $\hat{\delta}$'s toward their common mean. This shrinkage is explicit in the Bayesian estimator and implicit in our estimator, which is a weighted average of the $\hat{\delta}$'s across all managers (see equation 8). The bias of $\hat{\delta}^*$ is bigger than that of $\hat{\delta}^B$, and the (unreported) bias of the iterated estimator $\hat{\delta}^{*i}$ slightly exceeds the bias of $\hat{\delta}^*$. The changes estimator $\hat{\delta}^{**}$ is also biased, but unlike $\hat{\delta}^*$, it is biased even on average. This follows by design – while both $\hat{\delta}^*$ and $\hat{\delta}^{**}$ are weighted averages of the $\hat{\delta}$'s, the weights used in $\hat{\delta}^*$ sum to one, whereas those in $\hat{\delta}^{**}$ sum to zero, as shown earlier.

⁹Calculating the MSEs in Table 1 would not be sensible, as the units of γ and δ are different.

The bias with respect to δ present in our estimators suggests the need for caution when using these measures to estimate δ . It seems reasonable to use the levels estimator $\hat{\delta}^*$ for such a task, with similar costs and benefits as when using the Bayesian estimator, but the changes estimator $\hat{\delta}^{**}$ was not designed to capture δ even on average. More importantly, the bias does not compromise the success of our estimators in ranking managers by their δ or γ , as we saw in Tables 1 and 2. The simulation evidence reveals that our performance measures should be particularly useful in applications that involve such ranking, and one such application is presented in the following section.

The simulation results are robust to changes in the variance of stock alphas, σ_α^2 , as well as in the variance of unexpected stock returns, σ_e^2 . We also conducted the simulations in a setting in which each manager ranks stocks by their signals and invests equally in the top 25% of all stocks, instead of using a mean-variance criterion. This alternative setting leads to very similar results. Finally, note that we deliberately choose relatively small numbers of managers for the tables, $M = 10$ to 300, to be able to show some rare examples in which our measures are not clearly preferred to the standard ones. Since the usefulness of our measures grows with M , our measures are likely to be even more successful in ranking managers in practice as well as in our empirical analysis, where managers are counted in thousands.

Our measures have one other useful property that is easy to verify by simulation. Imagine a world with a large number of managers, all of whom exhibit no skill ($\delta = 0$) and whose holdings overlap sufficiently. The fact that there is no skill in this world would be hard to detect by $\hat{\delta}$, because the managers' $\hat{\delta}$'s would be dispersed around zero due to return noise. However, $\hat{\delta}^*$ and $\hat{\delta}^{**}$ should be zero for all managers, because the manager-specific noise is diversified away by averaging across many $\hat{\delta}$'s.

III Empirical Analysis

In this section, our skill measures are applied to evaluate the performance of a large sample of the U.S. equity mutual funds. After describing the data and our empirical methodology, we compare our measures with the standard measures in terms of their precision. We then describe the distribution of our measures and their significance across funds, and the persistence in mutual fund performance is investigated last.

III.1 Data and Methodology

Data on returns, prices, and shares outstanding of all NYSE, AMEX, and NASDAQ stocks are extracted from the monthly stock file provided by the Center for Research in Security Prices (CRSP). A firm's simple stock return is computed as an annual value-weighted return on the firm's common stock issues (typically one). Delisting returns are included when available on CRSP. If a firm is delisted but the delisting return is missing, we investigate the reason for disappearance. If the delisting is performance-related, we assume a -30% delisting return, following Shumway (1997); otherwise a zero delisting return is assumed. Quarterly stock returns are compounded from monthly returns.¹⁰ Market equity, measured as the combined value of all common stock classes outstanding, is taken from CRSP at the end of each month.

Mutual funds are required to file holdings reports with the SEC twice a year, but most funds publicly disclose their portfolio holdings on a quarterly basis. Thomson Financial collects and sells this data, which is currently available for the period from the end of 1975Q1 to the end of 2000Q3. The data is commonly known as the Spectrum data, since it used to be collected by CDA/Spectrum prior to their purchase by Thomson.¹¹ The Spectrum mutual fund holdings file contains four columns: date, stock identifier CUSIP, fund identifier, and the number of shares of the given stock held by the given fund on the given date. All dates are quarter-ends (3/31, 6/30, 9/30, or 12/31). Firms with no SPECTRUM data are recorded as having zero mutual fund ownership. We match each CUSIP to a CRSP PERMNO, the permanent number CRSP assigns to that security. Holdings associated with CUSIPs for which we found no associated PERMNO are ignored; these account for a very small fraction of fund holdings. Some companies have multiple equity securities associated with them. CRSP uniquely identifies each firm with a permanent company number, PERMCO. We value-weight returns and mutual fund holding percentages of the different share classes (PERMNOs) associated with each PERMCO, producing one return and one institutional ownership percentage associated with each firm.

Monthly fund returns are computed from the monthly returns on the stocks held

¹⁰If any of the two monthly stock returns following a valid return at the beginning of the quarter is missing, it is assumed to be zero when computing the cumulative stock return for this quarter.

¹¹See Wermers (1999) for detailed information regarding the construction of the database. This data is free of survival bias, as noted for example by Daniel, Grinblatt, Titman, and Wermers (1997).

by the fund at the most recent quarter-end, assuming that the fund follows a buy-and-hold strategy between the quarter-ends and that the quarterly rebalancing into the reported holdings is costless.¹² The resulting fund returns differ from the actual net fund returns in that they do not have expenses, management fees, and transaction costs subtracted from them, and that they contain only the equity portion of the fund return. These differences make our fund return series particularly well suited for an analysis of the funds' stock selection ability. This advantage comes at the cost of ignoring any intraquarter trading by the fund, including any potential window-dressing activity, as well as any ability of the fund to time the market by cleverly adjusting the fraction of its cash holdings.

One limitation of our data is that changes in holdings are observed only on a quarterly basis. If Fund 2 mimics the trades of Fund 1 with a short delay, we are unable to discern such a pattern unless the delay appears in the funds' quarterly holdings. Both funds are likely to be assigned similar performance by our measures, which seems appropriate only if Fund 1 is unable to profit on its trades before those trades are mimicked by Fund 2. Useful evidence on the importance of this issue is provided by Chen, Jegadeesh, and Wermers (2000), who find that stocks that experience net purchases by mutual funds have higher subsequent returns than stocks that experience net sales. The authors find that the abnormal performance following the funds' aggregate trades lasts for about a year. Most of the profits from fund trades are therefore not short-lived enough to confuse our performance measures. Also note that we do not separate the identity of fund managers from the funds they manage. Baks (2002) constructs a unique database of fund manager returns and characteristics by tracking the managers' career moves from one fund to another, and uses his database to examine fund manager performance.

At the end of each quarter in our sample, we compute various performance measures for each fund. First, we compute three versions of the traditional measure alpha, denoted earlier as $\hat{\delta}$, by regressing fund returns in excess of the risk-free rate on the benchmarks implied by three different pricing models. The CAPM alpha is calculated with respect to the market benchmark, as implied by the capital asset pricing model (CAPM), the Fama-French alpha is calculated with respect to the market, size, and value benchmarks, following Fama and French (1993), and the four-factor alpha is calculated with respect to

¹²This approach was first adopted by Grinblatt and Titman (1989).

the market, size, value, and momentum benchmarks, following Carhart (1997).¹³ Using any of the three $\hat{\delta}$'s as a reference measure, we then calculate the estimators of our measures, $\hat{\delta}^*$ and $\hat{\delta}^{**}$, from equations (4) and (22), as well as their iterated versions in equations (40) and (41).

III.2 How Precise Are the Proposed Performance Measures?

For all 3,281 funds that had existed for two or more years at the end of our holdings sample (September 30, 2000), we estimate all performance measures and their precisions using the complete return history for each fund. Two years of data are required to eliminate extremely noisy performance estimates.¹⁴ The precision of any estimator is assessed by the inverse of the estimator's squared standard error. The standard error of the traditional measure $\hat{\delta}$ is obtained from the usual regression. To compute the precisions of our measures $\hat{\delta}^*$ and $\hat{\delta}^{**}$ from equations (11) and (29), one needs to know Z , C , and Ω . The matrices Z and C are computed from fund holdings data, and Ω , the covariance matrix of $\hat{\delta}$, is computed following a procedure described in the appendix. The standard errors of $\hat{\delta}^*$ and $\hat{\delta}^{**}$ for each fund are given by the square roots of the diagonal elements of $Z\Omega Z'$ and $C\Omega C'$. For the iterated measures $\hat{\delta}^{*i}$ and $\hat{\delta}^{**i}$, the matrices Z and C are simply replaced by the products ZZ and CC .

Table 3 compares the precision of $\hat{\delta}$ with the precision of our estimators. The three panels correspond to three different definitions of the traditional measure alpha (CAPM, Fama-French, Carhart), which also serves as the reference measure for our measures. The first two rows of each panel report the typical ratios of the squared standard errors of our estimators to the squared standard error of $\hat{\delta}$. Both the ratio of the medians and the median ratio are reported. First, consider the levels estimator $\hat{\delta}^*$. For the CAPM alphas, the ratio of the medians is 0.22 and the median ratio is 0.24, which means that $\hat{\delta}^*$ is typically about four to five times more precise than $\hat{\delta}$. The precision ratio is below one for 93 percent of all funds. For the Fama-French alphas, the ratio of the medians

¹³All benchmark returns were obtained from Kenneth French's website.

¹⁴While this requirement induces some survival bias, this bias does not appear important. The average CAPM alpha of the funds eliminated from the sample is in fact slightly higher than the average alpha of the funds in our sample (-0.037 versus -0.054 percent per month), so excluding the short-history funds does not artificially increase the average fund performance. Also, the average CAPM alpha of the funds that died during the last 12 months of our sample is 0.038 percent per month, which suggests that our focus on the funds that are alive at the end of our sample does not introduce any survival bias either.

is 0.13 and the median ratio is 0.17, suggesting that $\hat{\delta}^*$ is on average about six times more precise than $\hat{\delta}$. Our estimator is more precise for 97 percent of all funds. For the four-factor alphas, $\hat{\delta}^*$ is more precise for 98 percent of all funds, and it is about six to eight times more precise for the median fund. As we add more benchmarks, the relative precision of $\hat{\delta}^*$ increases because $\hat{\delta}$'s become less correlated across funds, as discussed earlier. In all cases, iterating further increases the precision gains. The message from the table is clear – our holdings-based performance measure is estimated with much higher precision than the traditional measure.

The impressive precision of our levels estimator is not surprising. A typical fund holds dozens of stocks, and a typical stock is held by dozens of funds. Each additional holder improves the precision of the estimate of stock quality, and each additional stock makes fund performance estimation more precise. Nonetheless, Table 3 shows that, for a small minority of funds, $\hat{\delta}^*$ is less precise than $\hat{\delta}$. This happens especially for funds with long return histories that hold some stocks that are not extensively held by many other managers. If the fund's history spans several decades, its $\hat{\delta}$ is more precise than the $\hat{\delta}$'s of most other funds. Since $\hat{\delta}^*$ is a weighted average of many $\hat{\delta}$'s, it brings in the substantial sampling error contained in the $\hat{\delta}$'s of the shorter-lived funds, and this additional error can sometimes be large enough to make our measure less precise than $\hat{\delta}$. However, Table 3 makes it clear that this error is generally outweighed by the diversification effect of the averaging, making our measure more precise for a vast majority of funds. We also confirm that funds with shorter track records score bigger precision gains. The correlations between the length of the fund's return history and the precision ratios considered earlier are high, ranging from 0.45 to 0.56 across the three models.

The results for the changes estimator $\hat{\delta}^{**}$ are similar to those for $\hat{\delta}^*$. Our measure is more precise for 90 to 92 percent of all funds, and the precision increase is about sevenfold on average. Iterating further increases the average gain in precision, though the proportion of funds that experience precision gains does not change much. Shorter-history funds again experience bigger precision gains; the correlations between history length and the precision ratios range from 0.19 to 0.31 across the three models.

Of course, the standard error of a given estimator reflects the estimator's precision in estimating its own population value. While $\hat{\delta}$ estimates δ , $\hat{\delta}^*$ and $\hat{\delta}^{**}$ estimate their own population values δ^* and δ^{**} . Since the latter values are averages of δ 's, one might

expect them to be estimated more precisely. The usefulness of Table 3 consists mainly in quantifying these anticipated precision gains in mutual fund data.

III.3 How Many Funds Possess Significant Ability?

While the previous subsection analyzes the precision of our estimates, this subsection focuses on the estimates themselves, with an emphasis on their significance. We consider the performance estimates of the same 3,281 funds as before at the end of our sample. First, note that all measures are highly correlated with each other. The cross-fund correlation between $\hat{\delta}$ and $\hat{\delta}^*$ ranges from 0.64 to 0.70 across the three benchmark models, while the correlation between $\hat{\delta}$ and $\hat{\delta}^{**}$ ranges from 0.48 to 0.57. The correlations between our measures and their iterated counterparts exceed 0.91 in all cases.

Table 4 reports some summary statistics for $\hat{\delta}$, $\hat{\delta}^*$, $\hat{\delta}^{**}$, $\hat{\delta}^{*i}$, and $\hat{\delta}^{**i}$ across funds. To begin with the usual measure, about half of all funds have positive $\hat{\delta}$. The proportion of the positive $\hat{\delta}$'s is 0.50 for the CAPM, 0.46 for Fama-French, and 0.60 for the four-factor model. The distribution of $\hat{\delta}$ is quite symmetric across funds; while there is mild positive skewness, the mean exceeds the median in some cases but not in others. Estimates are considered significant if they are more than two standard errors from zero, i.e. if the absolute value of their t -statistic exceeds two. The proportions of funds with positive and negative t -statistics are similar: 0.03 vs 0.03 for the CAPM, 0.08 vs 0.07 for Fama-French, and 0.07 vs 0.02 for the four-factor model. We conclude that both the $\hat{\delta}$'s and their significance are distributed approximately symmetrically around zero.

In contrast, our levels measure $\hat{\delta}^*$ is much less symmetric. This measure has the same mean as $\hat{\delta}$, by construction, but less than half of its cross-fund standard deviation, due to averaging. Unlike $\hat{\delta}$, $\hat{\delta}^*$ is negatively skewed, which is easy to see from Figure 2.¹⁵ The distributions of $\hat{\delta}^*$ have fat left tails, revealing the presence of some “loser” funds with $\hat{\delta}^*$'s as low as minus one percent per month, whose existence is masked in the symmetric distribution of $\hat{\delta}$. Due to the negative skewness, $\hat{\delta}^*$'s median always exceeds the mean and $\hat{\delta}^*$ is positive for most funds – the fraction of funds with $\hat{\delta}^* > 0$ is 0.61 for the CAPM, 0.56 for Fama-French, and 0.83 for the four-factor model.

¹⁵The mildly positive skewness statistic in Panel C is a fluke driven by a small number of tail observations, because the plot of $\hat{\delta}^*$ in Figure 2 reveals obvious negative skewness. The median exceeds the mean and 67 percent of all $\hat{\delta}^*$ estimates are above the mean, compared to only 46 percent for $\hat{\delta}$.

To explain the negative skewness of $\hat{\delta}^*$, we introduce the concept of a “cousin”. The cousin of a given fund A is any fund B whose holdings overlap with those of A. Put differently, B is a cousin of A if there is at least one stock that is held by both A and B. For the 3,281 funds at the end of our sample, the number of cousins ranges from two to 3,210, with the average of 1,855 and the median of 2,106.

Interestingly, funds with higher $\hat{\delta}$ tend to have more cousins. The rank correlation between $\hat{\delta}$ and the number of cousins is 0.31 for the CAPM, 0.27 for Fama-French, and 0.16 for the four-factor model. This correlation might in part be due to the herding behavior of funds trying to mimic the holdings of funds with good track records. Since funds with fewer cousins are more likely to hold stocks that are held mostly by individuals, the positive correlation is also consistent with the finding of Cohen, Gompers, and Vuolteenaho (2002) that institutions as a group outperform individuals, albeit by a small margin. In any event, recall that a fund’s $\hat{\delta}^*$ is a weighted average of the $\hat{\delta}$ ’s of the fund’s cousins. Since the high- $\hat{\delta}$ funds have more cousins, the high $\hat{\delta}$ ’s are shrunk more toward the common mean than the low $\hat{\delta}$ ’s, creating negative skewness in $\hat{\delta}^*$.

The asymmetry in the significance of $\hat{\delta}^*$ is even more pronounced than the asymmetry in $\hat{\delta}^*$ itself. This is clear from the histograms of the t -statistics for $\hat{\delta}^*$, plotted in Figure 3, as well as from Table 4. While funds with significantly negative $\hat{\delta}^*$ account for less than one percent of all funds in all three models, the proportion of funds with significantly positive $\hat{\delta}^*$ is 0.05 for the CAPM, 0.19 for Fama-French, and 0.33 for the four-factor model. Under four benchmarks, only three funds exhibit significantly negative ability, but 1,076 funds (one out of three) exhibit significantly positive ability!

The key to understanding the striking asymmetry in the significance of the funds’ ability is the earlier observation that better-performing funds tend to have more cousins. Funds with higher $\hat{\delta}$ ’s have more cousins, funds with more cousins have $\hat{\delta}$ ’s that are better diversified across $\hat{\delta}$ ’s, and better diversified $\hat{\delta}$ ’s are generally more precise. The result is a strong negative correlation between the $\hat{\delta}$ ’s and their standard errors. This correlation ranges from -0.51 to -0.25 across the three models, as shown in Table 4. To construct the t -statistics, high $\hat{\delta}$ ’s are divided by their low standard errors, and low $\hat{\delta}$ ’s are divided by their high standard errors, which explains the asymmetry. Most of the left-tail loser funds from Figure 2 have insignificant $\hat{\delta}^*$ ’s due to their shortage of cousins, which translates into relatively large standard errors.

Note that the asymmetry in significance of $\hat{\delta}^*$ grows as we move from Panel A to Panel C. Under the CAPM benchmark, the number of funds with significantly positive ability is 14 times higher than the number of funds with significantly negative ability; this ratio grows to 22 under the Fama-French benchmarks, and to a staggering 359 under the four benchmarks in Panel C. This pattern is due to the rightward shift of the distribution of $\hat{\delta}^*$ (the average $\hat{\delta}^*$ increases from -0.054 in Panel A to -0.015 in Panel B to 0.112 in Panel C), and to some extent also due to the decline in the standard errors resulting from adding benchmarks that capture some common variation in returns.

In contrast to $\hat{\delta}^*$, the changes measure $\hat{\delta}^{**}$ does not exhibit much asymmetry around zero. Across the three models, 44 to 46 percent of all funds have positive $\hat{\delta}^{**}$. Funds with significantly negative $\hat{\delta}^{**}$ in fact slightly outnumber their positive counterparts, mostly due to the negative mean and median of $\hat{\delta}^{**}$. Also, the correlation between the estimates and their standard errors is small, unlike before. Iterating the changes measure into $\hat{\delta}^{**i}$ produces results similar to $\hat{\delta}^{**}$. The iterated levels measure $\hat{\delta}^{*i}$ has many features similar to its ancestor $\hat{\delta}^*$, such as strong asymmetry in significance (Panels B and C), but an exception occurs in Panel A, where no funds are inferred to have significantly positive skill. Compared to $\hat{\delta}^*$, $\hat{\delta}^{*i}$'s have the same negative mean but much less dispersion, which makes it harder for the positive $\hat{\delta}^{*i}$'s (which account for almost two thirds of all $\hat{\delta}^{*i}$'s) to be inferred significant.

In summary, the usual performance measure as well as our changes measure are distributed approximately symmetrically around zero across funds. According to our levels measure, though, funds with significantly positive ability before costs and fees vastly outnumber funds with significantly negative ability at the end of our sample.

III.4 Does Performance Persist?

III.4.1 Predicting Future Fund Returns

At the end of each quarter, funds are sorted into decile portfolios by various measures of past performance: $\hat{\delta}$, $\hat{\delta}^*$, and $\hat{\delta}^{**}$. Each measure is calculated three times, once for each benchmark model (CAPM, Fama-French, Carhart). The returns on the decile portfolios are calculated over the three months following the portfolio formation, equal-weighting

the funds within each decile. Since the deciles are rebalanced each quarter, the three-month return series is linked across quarters to form a monthly series of returns on each decile portfolio, covering the period April 1977 through December 2000. The beginning of the sample period is determined by our requirement that funds have at least two years of past data and by the fact that the holdings data become available at the end of 1975Q1. One more quarter of data is required upfront to construct changes in holdings, so the sample period for the changes estimator $\hat{\delta}^{**}$ begins in July 1977.

Table 5 reports the post-ranking alphas of the decile portfolios and of the “10-1” portfolio, constructed by going long the top decile of the best past performers and short the bottom decile. The alpha is defined with respect to the same benchmarks as the reference measure. That is, when the reference measure is say the CAPM alpha, the table reports the post-ranking CAPM alphas of the portfolios sorted on the past CAPM $\hat{\delta}$'s as well as on our measures that use the CAPM alpha as a reference measure.

Panel A of Table 5 reports results for funds sorted by their performance over the past 12 months. First, consider sorting funds on the traditional measure $\hat{\delta}$. When funds are sorted on their CAPM $\hat{\delta}$'s, the top decile earns a CAPM alpha of 3.8% per year ($t = 2.08$) and the bottom decile earns -4.5% per year ($t = -2.48$), which yields the 10-1 portfolio's CAPM alpha of 8.3% per year ($t = 2.99$). Strong persistence is preserved also when funds are sorted on the Fama-French $\hat{\delta}$'s; the 10-1 CAPM alpha is then 7.2% ($t = 3.52$). However, the 10-1 alpha drops to an insignificant 1.4% ($t = 0.73$) when the momentum benchmark is included. This result is in line with the evidence of Carhart (1997) and Wermers (1997), who argue that some apparent persistence in performance is due simply to momentum in stock returns. The argument is that, due to momentum, managers who happen to hold mostly stocks that performed well (poorly) over the past year are likely to do well (poorly) also over the following year, even in the absence of any rebalancing on their part. It seems hard to argue that sitting on one's laurels and doing nothing is a managerial skill that should be given credit.

Similar results are obtained when funds are sorted on our performance measures. Based on $\hat{\delta}^*$, the 10-1 CAPM alpha is 8.6% per year ($t = 2.57$), and the 10-1 Fama-French alpha is 8.3% ($t = 3.22$). After accounting for momentum, however, the difference again dips below significance, to 1.6% ($t = 0.65$). For $\hat{\delta}^{**}$, the 10-1 alphas are 4.4%, 3.7%, and 1.5% across the three models, with t -statistics of 3.78, 4.06, and 1.66, respectively.

When funds are sorted by performance estimated over the past 24 months (Panel B) or over their entire past history (Panel C), persistence is even weaker – the 10-1 CAPM and Fama-French alphas are positive for all three measures but never significant. Adding insult to injury, the 10-1 four-factor alphas are in fact insignificantly negative. The results based on the iterated measures $\hat{\delta}^{*i}$ and $\hat{\delta}^{**i}$, not reported to save space, lead to the same conclusions. (This statement applies to the following subsection as well.)

Significant persistence seems hard to find, even with our measures, which seem better suited for this task than $\hat{\delta}$. It is possible, though, that realized fund returns are too noisy to make persistence detectable by predicting future returns. Instead, the next subsection uses our skill measures to predict their own future values.

III.4.2 Predicting Future Skill

Each year between 1976 and 2000, at the end of the third quarter¹⁶, funds are sorted by $\hat{\delta}$, $\hat{\delta}^*$, and $\hat{\delta}^{**}$, estimated using return data over the past year. For each measure and each year, we compute the correlation between the fund ranking in that year and the ranking one year earlier. The time series of the resulting 24 rank correlations for each measure and each benchmark model is plotted in Figure 5. The average rank correlations and their time-series t -statistics are reported in Panel A of Table 6.

First, consider the usual measure $\hat{\delta}$. When funds are sorted on their usual CAPM alphas, the average rank correlation is significantly positive at 0.11 ($t = 3.00$), revealing significant persistence in market-adjusted performance. However, when funds are sorted on their Fama-French alphas, the average rank correlation drops to 0.04 ($t = 1.30$), and a sort on four-factor alphas produces further decline to 0.03 ($t = 1.68$).

Our measures also reveal weak persistence. With the CAPM alpha as the reference measure, both $\hat{\delta}^*$ and $\hat{\delta}^{**}$ display significant persistence; the rank correlations are 0.14 ($t = 2.58$) and 0.04 ($t = 2.55$), respectively. But the significance fades away when alternative reference measures are used. After accounting for momentum in stock returns, for example, the rank correlations drop to 0.03 ($t = 0.81$) and 0.02 ($t = 1.79$).

¹⁶Since our holdings sample ends in 2000Q3 and since the number of funds grows over time, choosing the third quarter maximizes the number of funds used in the analysis.

We also evaluate persistence at a two-year frequency. Every two years between 1979 and 2000, funds are sorted by their skill estimated over the previous two years, and correlations are computed between the rankings in adjacent two-year periods. Panel B of Table 6 reports these rank correlations, averaged across all 11 pairs of two-year periods. Persistence is generally weaker than in Panel A. The correlations range from 0.01 to 0.06 across all models and measures, but none of them is statistically significant. Given our simulation evidence that $\hat{\delta}^*$ and $\hat{\delta}^{**}$ are more accurate than $\hat{\delta}$ in ranking managers, the evidence presented in Table 6 reinforces the notion that mutual fund performance exhibits little persistence.

IV Conclusion

This paper proposes new performance measures that exploit the information contained in the similarity of a manager’s holdings (or changes in holdings) to those of managers who have performed well, and in their distinctiveness from those of managers who have performed poorly. These performance measures use historical returns and holdings of many funds to evaluate the performance of a single fund.¹⁷ As a result, these measures are typically about four to eight times more precise than the traditional return-based measures. While the usual measure as well as our trade-based measure are approximately symmetrically distributed around zero, at the end of our sample, funds with significantly positive skill greatly outnumber funds with significantly negative skill according to our holdings-based measure. In simulations, our measures are found to be particularly well suited for empirical applications that involve ranking managers. In one such application, we find only weak persistence in the before-cost performance of U.S. equity mutual funds after accounting for momentum in stock returns.

Our failure to find significant persistence does not necessarily mean that there is no persistent managerial skill. Small differences in ability across managers may not be detectable given the large amount of noise in realized fund returns. For example, Kothari

¹⁷Some information pooling across funds takes place also in the Bayesian frameworks developed by Jones and Shanken (2002) and Stambaugh (2002), in which a fund’s alpha is related to the alphas of other funds through a link in the prior. The techniques as well as the objectives of these papers are very different from ours. Jones and Shanken compare inferences about alphas with and without imposing their prior dependence, while Stambaugh focuses on inference about surviving funds. Neither study considers similarities in fund managers’ techniques and their relation to performance evaluation.

and Warner (2001) find that $\hat{\delta}$ has little ability to detect economically large magnitudes of abnormal fund performance. Although our measures seem superior to $\hat{\delta}$ in ranking managers, they may still not be powerful enough to find significant persistence. In fact, even if we were able to rank managers by their true δ 's, we might not be able to detect persistence in ex post returns at the conventional significance levels.

Importantly, fund returns may not be predictable even if some managers have superior ability. As explained by Berk and Green (2002), there should be no persistence in fund performance if investors compete for high returns and if managerial ability exhibits decreasing returns to scale. After good fund performance, rational investors infer that managerial skill is high, and new money flows into the fund until there is no abnormal expected future return. While the Berk-Green model implies no predictability in net fund returns, their intuition suggests that rational fund flows should weaken any persistence in gross returns as well. We find that some managers have superior skill before costs and fees but this skill exhibits only weak persistence over time, which is exactly what one should expect if the market for capital provision is competitive.

The basic idea in the paper is that managers who make similar investment decisions have similar skill. The proposed skill measures are designed to capture this idea in a simple way, but future research can extend these measures in various dimensions. For example, the measure of stock quality can be modified in some ad-hoc manner to give more weight to the $\hat{\delta}$'s with lower standard errors. As a broader example, useful information about similarities in investment decisions may also be contained in the funds' residual return correlations, if one is willing to assume that such similarities are stable over time. (No such assumption is needed when using holdings data, because similarity in holdings can be evaluated across stocks at any given point in time.) Finally, our measures rely on current holdings or recent changes in holdings. The earlier holdings are relevant only to the extent that they determine the manager's own track record, which is only one of many track records reflected in our measures. It seems reasonable to infer current skill from current rather than past decisions, especially if we intend to forecast the manager's future performance. Nonetheless, it might be interesting to design performance measures that incorporate also similarities among the managers' historical holdings. Future research can also pursue a variety of applications of our measures.

Table 1
Simulation Evidence on Various Skill Measures Relative to True Skill γ

Random samples of returns on N stocks and signals of M managers are simulated as described in Section 2, with q denoting the weight on the common signal (which makes managers' portfolios more alike). In each sample, we calculate eight different performance measures for each manager. Three of these measures are population versions of the traditional measure δ and of our levels and changes estimators, δ^* and δ^{**} . The remaining five measures are sample measures: $\hat{\delta}$ denotes the actual return on the manager's portfolio, $\hat{\delta}^*$ denotes our levels estimator, which exploits similarities in holdings across managers, $\hat{\delta}^{*i}$ denotes the iterated version of our levels estimator, $\hat{\delta}^{**}$ denotes our changes estimator, which exploits similarities in changes in holdings across managers, and $\hat{\delta}^{**i}$ denotes the iterated version of our changes estimator. All managers are ranked according to their performance estimated using the eight measures, and rank correlations with the ranking based on true skill γ are reported. All numbers are averaged across 10,000 simulations.

Rank Correlations with True Skill (γ)																
M	q=0								q=0.50							
	$\hat{\delta}$	$\hat{\delta}^*$	$\hat{\delta}^{*i}$	$\hat{\delta}^{**}$	$\hat{\delta}^{**i}$	δ	δ^*	δ^{**}	$\hat{\delta}$	$\hat{\delta}^*$	$\hat{\delta}^{*i}$	$\hat{\delta}^{**}$	$\hat{\delta}^{**i}$	δ	δ^*	δ^{**}
N=10																
10	0.30	0.30	0.28	0.30	0.28	0.59	0.55	0.58	0.16	0.15	0.13	0.15	0.13	0.33	0.30	0.32
50	0.33	0.38	0.39	0.39	0.42	0.63	0.63	0.68	0.17	0.18	0.18	0.18	0.18	0.36	0.35	0.39
100	0.32	0.39	0.42	0.40	0.45	0.63	0.64	0.69	0.18	0.19	0.19	0.20	0.20	0.37	0.36	0.40
300	0.33	0.40	0.44	0.42	0.47	0.64	0.64	0.70	0.18	0.20	0.21	0.20	0.21	0.37	0.37	0.41
N=50																
10	0.61	0.64	0.62	0.65	0.64	0.83	0.82	0.83	0.37	0.35	0.32	0.36	0.32	0.62	0.59	0.59
50	0.66	0.81	0.84	0.83	0.86	0.87	0.88	0.89	0.40	0.45	0.44	0.46	0.46	0.66	0.64	0.66
100	0.66	0.84	0.87	0.86	0.89	0.88	0.89	0.90	0.40	0.49	0.50	0.50	0.53	0.66	0.65	0.68
300	0.67	0.87	0.89	0.88	0.90	0.88	0.90	0.91	0.40	0.53	0.57	0.54	0.61	0.67	0.67	0.70
N=100																
10	0.74	0.78	0.77	0.78	0.78	0.89	0.88	0.89	0.49	0.48	0.45	0.49	0.46	0.73	0.71	0.72
50	0.78	0.91	0.92	0.92	0.93	0.93	0.93	0.94	0.53	0.61	0.62	0.62	0.64	0.78	0.76	0.78
100	0.79	0.93	0.94	0.94	0.94	0.93	0.94	0.95	0.53	0.66	0.68	0.67	0.71	0.78	0.78	0.80
300	0.79	0.94	0.94	0.95	0.95	0.94	0.94	0.95	0.54	0.71	0.75	0.73	0.78	0.79	0.80	0.82

Table 2
Simulation Evidence on Various Skill Measures Relative to Traditional Skill δ

Random samples of returns on N stocks and signals of M managers are simulated as described in Section 2, with q denoting the weight on the common signal (which makes managers' portfolios more alike). In each sample, we calculate eight different performance measures for each manager. Two of these measures are population versions of our levels and changes estimators, δ^* and δ^{**} . The remaining six measures are sample measures: $\hat{\delta}$ denotes the actual return on the manager's portfolio, $\hat{\delta}^B$ denotes the Bayesian estimator that shrinks $\hat{\delta}$ halfway toward the average of $\hat{\delta}$ across managers, $\hat{\delta}^*$ denotes our levels estimator, which exploits similarities in holdings across managers, $\hat{\delta}^{*i}$ denotes the iterated version of our levels estimator, $\hat{\delta}^{**}$ denotes our changes estimator, which exploits similarities in changes in holdings across managers, and $\hat{\delta}^{**i}$ denotes the iterated version of our changes estimator. All managers are ranked according to their performance estimated using the eight measures, and rank correlations with the ranking based on the traditional skill measure δ are reported in Panel A. Panel B reports averages across managers of the squared differences between the estimator and the true δ (times 100). All numbers are averaged across 10,000 simulations.

Panel A. Rank Correlations with Traditional Skill (δ)																
M	q=0								q=0.50							
	$\hat{\delta}$	$\hat{\delta}^B$	$\hat{\delta}^*$	$\hat{\delta}^{*i}$	$\hat{\delta}^{**}$	$\hat{\delta}^{**i}$	δ^*	δ^{**}	$\hat{\delta}$	$\hat{\delta}^B$	$\hat{\delta}^*$	$\hat{\delta}^{*i}$	$\hat{\delta}^{**}$	$\hat{\delta}^{**i}$	δ^*	δ^{**}
N=10																
10	0.46	0.46	0.46	0.43	0.43	0.40	0.91	0.86	0.42	0.42	0.39	0.35	0.35	0.31	0.89	0.81
50	0.50	0.50	0.57	0.59	0.53	0.56	0.97	0.92	0.45	0.45	0.47	0.45	0.41	0.40	0.96	0.89
100	0.50	0.50	0.58	0.62	0.55	0.59	0.98	0.93	0.46	0.46	0.50	0.51	0.44	0.45	0.98	0.91
300	0.50	0.50	0.61	0.66	0.57	0.63	0.99	0.94	0.46	0.46	0.51	0.54	0.45	0.47	0.99	0.92
N=50																
10	0.70	0.70	0.73	0.71	0.73	0.71	0.94	0.94	0.55	0.55	0.53	0.48	0.52	0.47	0.92	0.90
50	0.74	0.74	0.90	0.93	0.90	0.93	0.98	0.97	0.59	0.59	0.65	0.63	0.63	0.63	0.94	0.93
100	0.75	0.75	0.93	0.96	0.92	0.95	0.98	0.97	0.59	0.59	0.70	0.71	0.68	0.71	0.95	0.94
300	0.75	0.75	0.95	0.97	0.95	0.96	0.99	0.98	0.60	0.60	0.76	0.81	0.74	0.80	0.97	0.96
N=100																
10	0.80	0.80	0.83	0.82	0.84	0.83	0.96	0.96	0.62	0.62	0.61	0.58	0.61	0.57	0.94	0.93
50	0.84	0.84	0.96	0.97	0.96	0.97	0.98	0.98	0.67	0.67	0.75	0.75	0.75	0.76	0.95	0.95
100	0.84	0.84	0.97	0.98	0.97	0.98	0.99	0.98	0.67	0.67	0.81	0.83	0.80	0.83	0.96	0.95
300	0.85	0.85	0.98	0.99	0.98	0.98	0.99	0.99	0.68	0.68	0.87	0.91	0.86	0.91	0.97	0.96

Panel B. Mean Squared Errors																
M	q=0								q=0.50							
	$\hat{\delta}$	$\hat{\delta}^B$	$\hat{\delta}^*$	$\hat{\delta}^{*i}$	$\hat{\delta}^{**}$	$\hat{\delta}^{**i}$	δ^*	δ^{**}	$\hat{\delta}$	$\hat{\delta}^B$	$\hat{\delta}^*$	$\hat{\delta}^{*i}$	$\hat{\delta}^{**}$	$\hat{\delta}^{**i}$	δ^*	δ^{**}
N=10																
10	1.27	0.88	0.87	0.86	1.60	4.57	0.08	0.25	1.26	0.87	0.85	0.84	1.39	3.27	0.08	0.20
50	1.27	0.85	0.79	0.83	0.96	1.40	0.11	0.16	1.25	0.82	0.77	0.79	0.80	0.91	0.10	0.18
100	1.27	0.85	0.79	0.83	0.86	1.09	0.12	0.15	1.25	0.82	0.76	0.79	0.73	0.73	0.11	0.17
300	1.27	0.84	0.77	0.82	0.82	0.94	0.12	0.14	1.26	0.83	0.77	0.80	0.68	0.62	0.11	0.16
N=50																
10	0.24	0.20	0.21	0.25	0.19	0.27	0.07	0.09	0.24	0.18	0.18	0.19	0.22	0.29	0.04	0.15
50	0.24	0.19	0.22	0.27	0.14	0.23	0.10	0.11	0.24	0.17	0.18	0.19	0.28	0.40	0.05	0.27
100	0.24	0.19	0.23	0.27	0.14	0.24	0.11	0.11	0.24	0.17	0.18	0.19	0.27	0.40	0.06	0.27
300	0.24	0.19	0.23	0.27	0.14	0.25	0.11	0.12	0.24	0.17	0.18	0.19	0.28	0.41	0.06	0.27
N=100																
10	0.12	0.11	0.14	0.17	0.13	0.19	0.07	0.09	0.12	0.09	0.10	0.11	0.18	0.27	0.03	0.15
50	0.12	0.11	0.15	0.19	0.13	0.24	0.09	0.12	0.12	0.09	0.11	0.11	0.29	0.40	0.04	0.29
100	0.12	0.11	0.16	0.19	0.13	0.26	0.10	0.12	0.12	0.09	0.10	0.11	0.29	0.40	0.04	0.29
300	0.12	0.11	0.16	0.20	0.14	0.28	0.10	0.13	0.12	0.09	0.11	0.11	0.29	0.41	0.04	0.29

Table 3
Precision of Various Performance Measures

The table compares the precision of the traditional estimator $\hat{\delta}$ with the precision of our estimators $\hat{\delta}^*$, $\hat{\delta}^{**}$, and their iterated versions. All summary statistics are calculated across the 3,281 funds that had existed for two or more years at the end of our sample. The first row of each panel reports the ratio of the median squared standard error of our estimator to the median squared standard error of $\hat{\delta}$. The second row reports the median ratio of the squared standard error of our estimator to the squared standard error of $\hat{\delta}$. Row three shows what fraction of these ratios are below one. Rows four, five, and six report some plain summary statistics for the standard errors across funds, expressed in percent per month. The reference measure alpha, which coincides with the traditional measure $\hat{\delta}$, is defined differently across the three panels. The CAPM alpha is computed with respect to the market benchmark, the Fama-French alpha with respect to the market, size, and value benchmarks, following Fama and French (1993), and the four-factor alpha with respect to the Fama-French and momentum benchmarks, following Carhart (1997).

	$\hat{\delta}$	$\hat{\delta}^*$	$\hat{\delta}^{**}$	$\hat{\delta}^{*i}$	$\hat{\delta}^{**i}$
Panel A: CAPM Alpha as a Reference Measure					
Median(SE ²)/Median(SE($\hat{\delta}$) ²)		0.22	0.14	0.13	0.07
Median(SE ² /SE($\hat{\delta}$) ²)		0.24	0.15	0.16	0.09
Fraction of Ratios < 1		0.93	0.90	0.97	0.91
Mean SE	0.59	0.28	0.26	0.22	0.26
Median SE	0.42	0.20	0.16	0.15	0.11
Std SE	0.51	0.24	0.36	0.19	0.57
Panel B: Fama-French Alpha as a Reference Measure					
Median(SE ²)/Median(SE($\hat{\delta}$) ²)		0.13	0.12	0.09	0.07
Median(SE ² /SE($\hat{\delta}$) ²)		0.17	0.15	0.13	0.10
Fraction of Ratios < 1		0.97	0.92	0.98	0.91
Mean SE	0.48	0.19	0.22	0.16	0.23
Median SE	0.34	0.12	0.12	0.10	0.09
Std SE	0.45	0.20	0.33	0.16	0.53
Panel C: Four-Factor Alpha as a Reference Measure					
Median(SE ²)/Median(SE($\hat{\delta}$) ²)		0.12	0.12	0.08	0.07
Median(SE ² /SE($\hat{\delta}$) ²)		0.16	0.14	0.13	0.10
Fraction of Ratios < 1		0.98	0.92	0.98	0.91
Mean SE	0.51	0.19	0.23	0.16	0.24
Median SE	0.35	0.12	0.12	0.10	0.10
Std SE	0.47	0.21	0.35	0.16	0.56

Table 4
Summary Statistics and Significance for Various Performance Measures

The table summarizes the distributions of the estimators listed in the column headings. All summary statistics are calculated across the 3,281 funds that had existed for two or more years at the end of our sample. Each panel reports the proportions of all performance estimates that are positive (row 1), more than two standard errors above zero (row 2), and more than two standard errors below zero (row 3). The table also reports the estimates' rank correlations with the number of cousins (row 8) and the estimates' correlations with their own standard errors (row 9), as well as some plain summary statistics for the estimates, expressed in percent per month. The reference measure alpha, which coincides with the traditional measure $\hat{\delta}$, is defined differently across the three panels. The CAPM alpha is computed with respect to the market benchmark, the Fama-French alpha with respect to the market, size, and value benchmarks, following Fama and French (1993), and the four-factor alpha with respect to the Fama-French and momentum benchmarks, following Carhart (1997).

	$\hat{\delta}$	$\hat{\delta}^*$	$\hat{\delta}^{**}$	$\hat{\delta}^{*i}$	$\hat{\delta}^{**i}$
Panel A: CAPM Alpha as a Reference Measure					
Fraction of Estimates > 0	0.504	0.614	0.463	0.644	0.462
Fraction of Estimates > 2 SE	0.028	0.047	0.018	0.000	0.024
Fraction of Estimates < -2 SE	0.031	0.003	0.032	0.001	0.025
Mean	-0.054	-0.054	-0.031	-0.054	-0.024
Median	0.003	0.047	-0.012	0.051	-0.009
Standard deviation	0.827	0.359	0.467	0.281	0.636
Skewness	0.704	-2.067	-1.691	-2.885	-5.740
Correlation with Number of Cousins	0.312	0.541	0.081	0.555	0.036
Correlation with SE	0.068	-0.515	-0.010	-0.670	-0.048
Panel B: Fama-French Alpha as a Reference Measure					
Fraction of Estimates > 0	0.465	0.564	0.440	0.676	0.462
Fraction of Estimates > 2 SE	0.084	0.186	0.052	0.077	0.045
Fraction of Estimates < -2 SE	0.073	0.009	0.091	0.003	0.066
Mean	-0.015	-0.015	-0.036	-0.015	-0.026
Median	-0.035	0.033	-0.021	0.054	-0.010
Standard deviation	0.844	0.364	0.477	0.273	0.656
Skewness	0.908	-1.330	-1.826	-2.514	-7.163
Correlation with Number of Cousins	0.274	0.417	0.075	0.411	0.028
Correlation with SE	0.047	-0.508	0.002	-0.666	-0.054
Panel C: Four-Factor Alpha as a Reference Measure					
Fraction of Estimates > 0	0.603	0.827	0.447	0.851	0.470
Fraction of Estimates > 2 SE	0.074	0.328	0.033	0.300	0.034
Fraction of Estimates < -2 SE	0.015	0.001	0.045	0.001	0.039
Mean	0.112	0.112	-0.013	0.112	-0.012
Median	0.084	0.143	-0.015	0.153	-0.008
Standard deviation	0.760	0.282	0.529	0.213	0.767
Skewness	1.530	0.116	-1.366	-1.026	-8.534
Correlation with Number of Cousins	0.164	0.382	0.020	0.373	-0.014
Correlation with SE	0.185	-0.254	0.080	-0.398	-0.018

Table 5
Performance Persistence Measured by Alphas of Portfolios of Funds Sorted by
Various Measures of Past Performance

At the end of each quarter, funds are sorted into decile portfolios by various measures of past performance: $\hat{\delta}$, the OLS estimate of the fund's alpha, $\hat{\delta}^*$, our levels estimator, and $\hat{\delta}^{**}$, our changes estimator. The returns on the decile portfolios are calculated over the three months after portfolio formation, equal-weighting the funds within each decile. The three-month return series are linked across quarters to form a monthly series of returns on each decile portfolio. All performance measures are calculated using the data over the past 12 months (Panel A), 24 months (Panel B), and all available past data (Panel C). All portfolio return series end in December 2000; they begin in April 1977 for $\hat{\delta}$ and $\hat{\delta}^*$ and in July 1977 for $\hat{\delta}^{**}$. The table reports the OLS estimates of the deciles' full-period alphas (in percent per year), as well as their t -statistics (in parentheses). These alphas are defined in the same way as the reference measures $\hat{\delta}$, and they differ across the three panels. The CAPM alpha is computed with respect to the market benchmark, the Fama-French alpha with respect to the market, size, and value benchmarks, following Fama and French (1993), and the four-factor alpha with respect to the Fama-French and momentum benchmarks, following Carhart (1997).

	Decile of Funds Sorted by Past Performance										
	1	2	3	4	5	6	7	8	9	10	10-1
Panel A. Sorting Funds by Past 12-Month Performance											
CAPM Alpha as a Reference Measure											
$\hat{\delta}$	-4.48	-2.40	-1.39	-0.56	-0.25	0.14	0.95	1.70	2.20	3.83	8.31
	(-2.48)	(-2.24)	(-1.67)	(-0.88)	(-0.50)	(0.30)	(1.80)	(2.29)	(1.93)	(2.08)	(2.99)
$\hat{\delta}^*$	-5.07	-2.70	-1.86	-0.85	-0.11	0.71	1.23	1.98	2.84	3.50	8.56
	(-2.30)	(-2.04)	(-2.01)	(-1.24)	(-0.21)	(1.50)	(2.04)	(2.05)	(1.99)	(1.66)	(2.57)
$\hat{\delta}^{**}$	-2.18	-1.33	-0.35	-0.40	0.00	0.20	0.48	1.06	0.97	2.20	4.37
	(-1.83)	(-1.82)	(-0.62)	(-0.76)	(0.00)	(0.39)	(0.94)	(1.75)	(1.19)	(1.63)	(3.78)
Fama-French Alpha as a Reference Measure											
$\hat{\delta}$	-3.87	-2.12	-1.51	-0.34	-0.27	0.27	0.59	1.69	2.18	3.32	7.20
	(-2.34)	(-2.50)	(-2.55)	(-0.63)	(-0.56)	(0.59)	(1.29)	(3.34)	(3.58)	(2.73)	(3.52)
$\hat{\delta}^*$	-4.99	-2.63	-1.58	-1.08	-0.49	0.33	1.54	2.38	3.13	3.26	8.26
	(-2.48)	(-2.56)	(-1.92)	(-1.69)	(-0.91)	(0.70)	(2.81)	(3.67)	(3.93)	(2.20)	(3.22)
$\hat{\delta}^{**}$	-2.77	-0.25	-0.07	-0.01	0.13	0.33	0.41	1.15	0.78	0.88	3.65
	(-2.73)	(-0.42)	(-0.14)	(-0.02)	(0.27)	(0.67)	(0.84)	(2.23)	(1.29)	(0.89)	(4.06)
Four-Factor Alpha as a Reference Measure											
$\hat{\delta}$	-0.36	-0.56	-0.26	0.39	0.37	0.12	0.68	1.37	1.20	0.99	1.35
	(-0.23)	(-0.72)	(-0.44)	(0.70)	(0.76)	(0.24)	(1.44)	(2.63)	(1.91)	(0.82)	(0.73)
$\hat{\delta}^*$	0.08	-0.65	-0.58	-0.32	-0.09	0.31	0.32	1.04	2.13	1.67	1.59
	(0.04)	(-0.65)	(-0.69)	(-0.46)	(-0.17)	(0.65)	(0.59)	(1.64)	(2.30)	(1.09)	(0.65)
$\hat{\delta}^{**}$	-0.79	0.43	0.06	0.65	0.85	0.53	0.98	0.74	0.42	0.72	1.50
	(-0.82)	(0.75)	(0.12)	(1.24)	(1.62)	(1.02)	(1.90)	(1.32)	(0.65)	(0.67)	(1.66)

Table 5 - cont'd

	Decile of Funds Sorted by Past Performance										
	1	2	3	4	5	6	7	8	9	10	10-1
Panel B. Sorting Funds by Past 24-Month Performance											
CAPM Alpha as a Reference Measure											
$\hat{\delta}$	-1.16	-0.66	-0.55	0.08	0.13	0.63	0.78	1.35	1.17	1.60	2.77
	(-0.68)	(-0.62)	(-0.67)	(0.14)	(0.25)	(1.37)	(1.58)	(2.14)	(1.31)	(0.94)	(1.10)
$\hat{\delta}^*$	-1.06	-1.16	-0.55	-0.08	0.79	0.72	0.91	1.18	1.84	0.72	1.79
	(-0.49)	(-0.96)	(-0.57)	(-0.12)	(1.55)	(1.44)	(1.67)	(1.68)	(1.66)	(0.35)	(0.56)
$\hat{\delta}^{**}$	-0.84	0.01	-0.52	0.09	0.52	0.46	0.74	1.01	1.20	0.60	1.44
	(-0.74)	(0.02)	(-0.87)	(0.16)	(1.04)	(0.98)	(1.50)	(1.78)	(1.63)	(0.48)	(1.25)
Fama-French Alpha as a Reference Measure											
$\hat{\delta}$	-1.18	-1.18	-0.54	-0.58	-0.19	0.16	1.27	0.92	2.49	2.70	3.88
	(-0.74)	(-1.50)	(-0.92)	(-1.04)	(-0.37)	(0.34)	(2.51)	(1.74)	(3.97)	(2.31)	(1.91)
$\hat{\delta}^*$	-1.26	-1.24	-1.01	-1.28	-0.40	0.74	0.87	1.52	2.72	3.18	4.44
	(-0.66)	(-1.28)	(-1.22)	(-1.78)	(-0.72)	(1.53)	(1.62)	(2.32)	(3.32)	(2.17)	(1.74)
$\hat{\delta}^{**}$	-1.18	-0.36	0.10	0.08	0.46	0.47	0.82	1.34	1.21	0.75	1.92
	(-1.27)	(-0.62)	(0.19)	(0.15)	(0.90)	(1.01)	(1.67)	(2.54)	(2.20)	(0.79)	(1.99)
Four-Factor Alpha as a Reference Measure											
$\hat{\delta}$	1.65	0.00	0.39	0.51	0.60	0.69	0.80	1.27	1.19	0.25	-1.40
	(1.05)	(0.00)	(0.65)	(0.94)	(1.21)	(1.34)	(1.53)	(2.38)	(1.93)	(0.22)	(-0.76)
$\hat{\delta}^*$	1.27	0.87	0.42	0.17	0.38	0.68	0.71	1.14	1.32	0.30	-0.97
	(0.67)	(0.88)	(0.53)	(0.26)	(0.68)	(1.28)	(1.23)	(1.75)	(1.66)	(0.22)	(-0.40)
$\hat{\delta}^{**}$	0.28	0.57	0.61	0.43	0.82	0.79	1.00	1.29	1.27	0.07	-0.20
	(0.29)	(0.98)	(1.23)	(0.83)	(1.61)	(1.55)	(1.90)	(2.33)	(2.17)	(0.08)	(-0.23)
Panel C. Sorting Funds by Entire Past Performance Record											
CAPM Alpha as a Reference Measure											
$\hat{\delta}$	-1.69	-0.51	-0.09	0.39	0.91	0.91	1.08	1.00	0.99	0.31	2.00
	(-1.15)	(-0.56)	(-0.13)	(0.63)	(1.71)	(1.88)	(2.26)	(1.66)	(1.41)	(0.24)	(1.11)
$\hat{\delta}^*$	-2.18	-0.80	0.05	1.08	1.41	1.35	1.11	1.06	0.55	-0.25	1.94
	(-1.09)	(-0.67)	(0.06)	(1.30)	(2.23)	(2.47)	(2.09)	(1.67)	(0.56)	(-0.14)	(0.71)
$\hat{\delta}^{**}$	-0.32	-0.22	0.16	0.52	0.38	0.67	0.57	0.83	0.65	0.13	0.46
	(-0.28)	(-0.31)	(0.27)	(1.04)	(0.79)	(1.38)	(1.16)	(1.71)	(0.99)	(0.11)	(0.50)
Fama-French Alpha as a Reference Measure											
$\hat{\delta}$	-0.99	-0.51	-0.69	-0.01	0.25	0.23	1.18	0.35	1.66	2.27	3.26
	(-0.71)	(-0.67)	(-1.07)	(-0.01)	(0.46)	(0.48)	(2.77)	(0.73)	(3.13)	(2.54)	(1.99)
$\hat{\delta}^*$	-1.72	-0.91	-0.37	-0.27	-0.21	0.30	0.48	1.69	2.38	2.43	4.15
	(-0.94)	(-0.91)	(-0.49)	(-0.35)	(-0.34)	(0.56)	(0.93)	(2.78)	(3.03)	(1.83)	(1.76)
$\hat{\delta}^{**}$	-0.53	0.06	0.41	0.01	-0.01	0.54	1.00	0.89	0.77	0.61	1.14
	(-0.55)	(0.10)	(0.78)	(0.02)	(-0.02)	(1.12)	(2.23)	(1.77)	(1.43)	(0.67)	(1.43)
Four-Factor Alpha as a Reference Measure											
$\hat{\delta}$	0.89	0.37	0.42	0.23	0.27	0.72	1.27	0.71	1.42	0.84	-0.06
	(0.68)	(0.56)	(0.66)	(0.41)	(0.51)	(1.49)	(2.46)	(1.28)	(2.41)	(1.04)	(-0.04)
$\hat{\delta}^*$	1.00	-0.39	0.35	0.68	0.58	0.81	0.95	1.56	1.54	0.10	-0.90
	(0.55)	(-0.40)	(0.44)	(0.88)	(0.90)	(1.42)	(1.69)	(2.46)	(1.84)	(0.08)	(-0.40)
$\hat{\delta}^{**}$	0.23	0.30	0.42	0.56	0.82	1.06	1.06	1.08	0.94	0.64	0.42
	(0.23)	(0.53)	(0.86)	(1.09)	(1.63)	(2.10)	(2.06)	(2.10)	(1.54)	(0.71)	(0.58)

Table 6
Performance Persistence Measured by Serial Rank Correlations

The table reports the time series averages of the rank correlations across funds between performance measures estimated at the end of a given year and 12 months (Panel A) or 24 months (Panel B) earlier. The t -statistics based on the time series (1977 through 2000 in Panel A, 1980 through 2000 in Panel B) are in parentheses. The reference measure alpha coincides with the traditional measure $\hat{\delta}$. The CAPM alpha is computed with respect to the market benchmark, the Fama-French alpha with respect to the market, size, and value benchmarks, following Fama and French (1993), and the four-factor alpha with respect to the Fama-French and momentum benchmarks, following Carhart (1997).

Reference Measure	$\hat{\delta}$	$\hat{\delta}^*$	$\hat{\delta}^{**}$
Panel A: 12-Month Frequency			
CAPM Alpha	0.11 (3.00)	0.14 (2.58)	0.04 (2.55)
Fama-French Alpha	0.04 (1.30)	0.05 (1.18)	0.03 (1.81)
Four-Factor Alpha	0.03 (1.68)	0.03 (0.81)	0.02 (1.79)
Panel B: 24-Month Frequency			
CAPM Alpha	0.05 (0.82)	0.02 (0.24)	0.02 (1.33)
Fama-French Alpha	0.06 (1.99)	0.06 (0.91)	0.02 (1.25)
Four-Factor Alpha	0.04 (1.17)	0.03 (0.53)	0.01 (0.33)

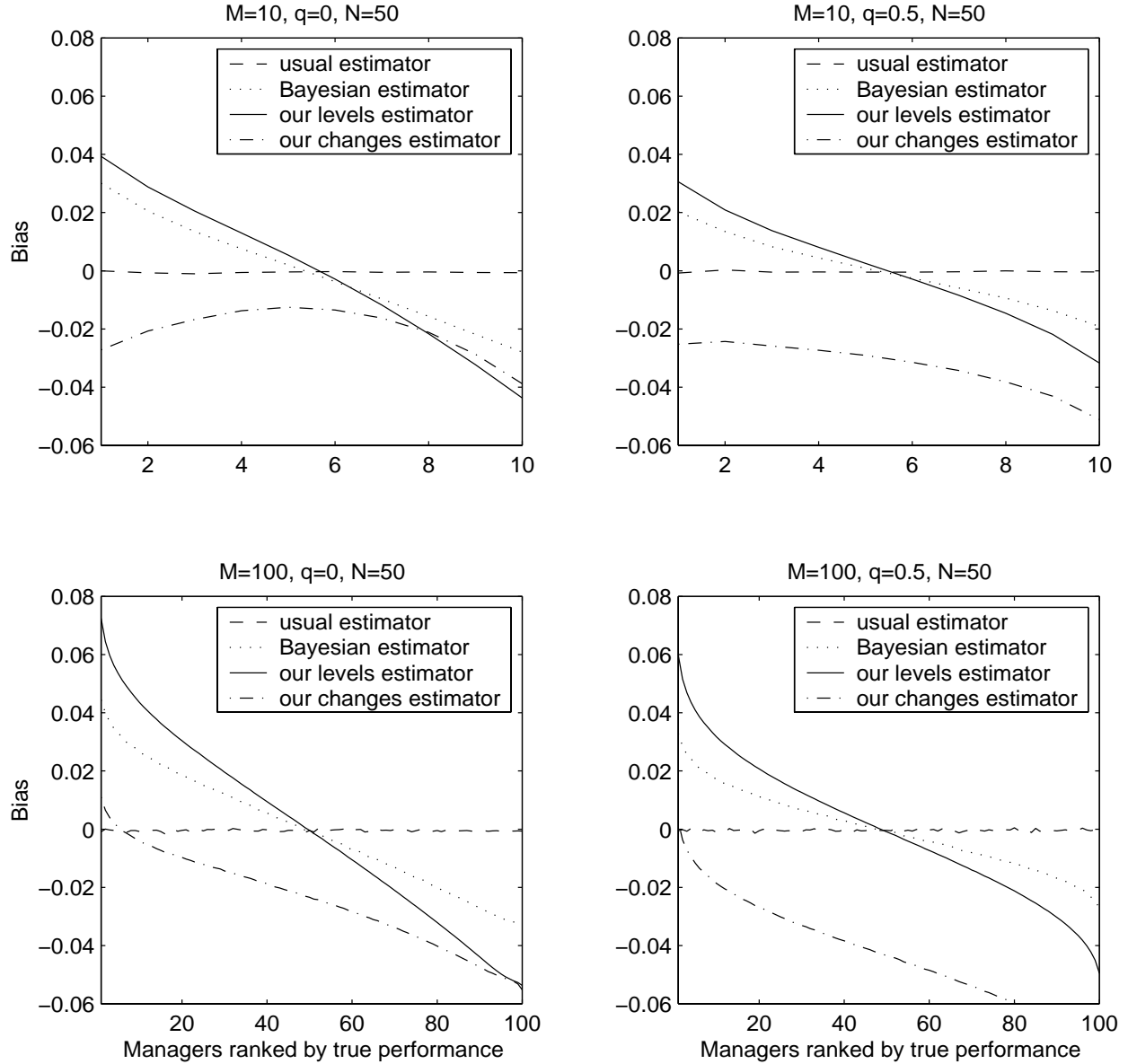


Figure 1. Bias of alternative performance estimators with respect to the traditional performance measure.

The figure plots the biases of four alternative performance estimators for managers ranked by the true value of the traditional performance measure δ . In each simulated sample with M (10 or 100) managers, $N = 50$ stocks, and the weight q (0 or 0.5) on the common signal, managers are ranked in ascending order by their δ . Denote the true performance of the m -th ranked manager by δ_m . For each estimator $\tilde{\delta}$ and each rank m ($m = 1, \dots, M$), the bias is computed by averaging $\tilde{\delta}_m - \delta_m$ across 10,000 simulated samples. Four estimators are considered: the usual estimator $\hat{\delta}$, the Bayesian estimator $\hat{\delta}^B$, our levels estimator $\hat{\delta}^*$, and our changes estimator $\hat{\delta}^{**}$.

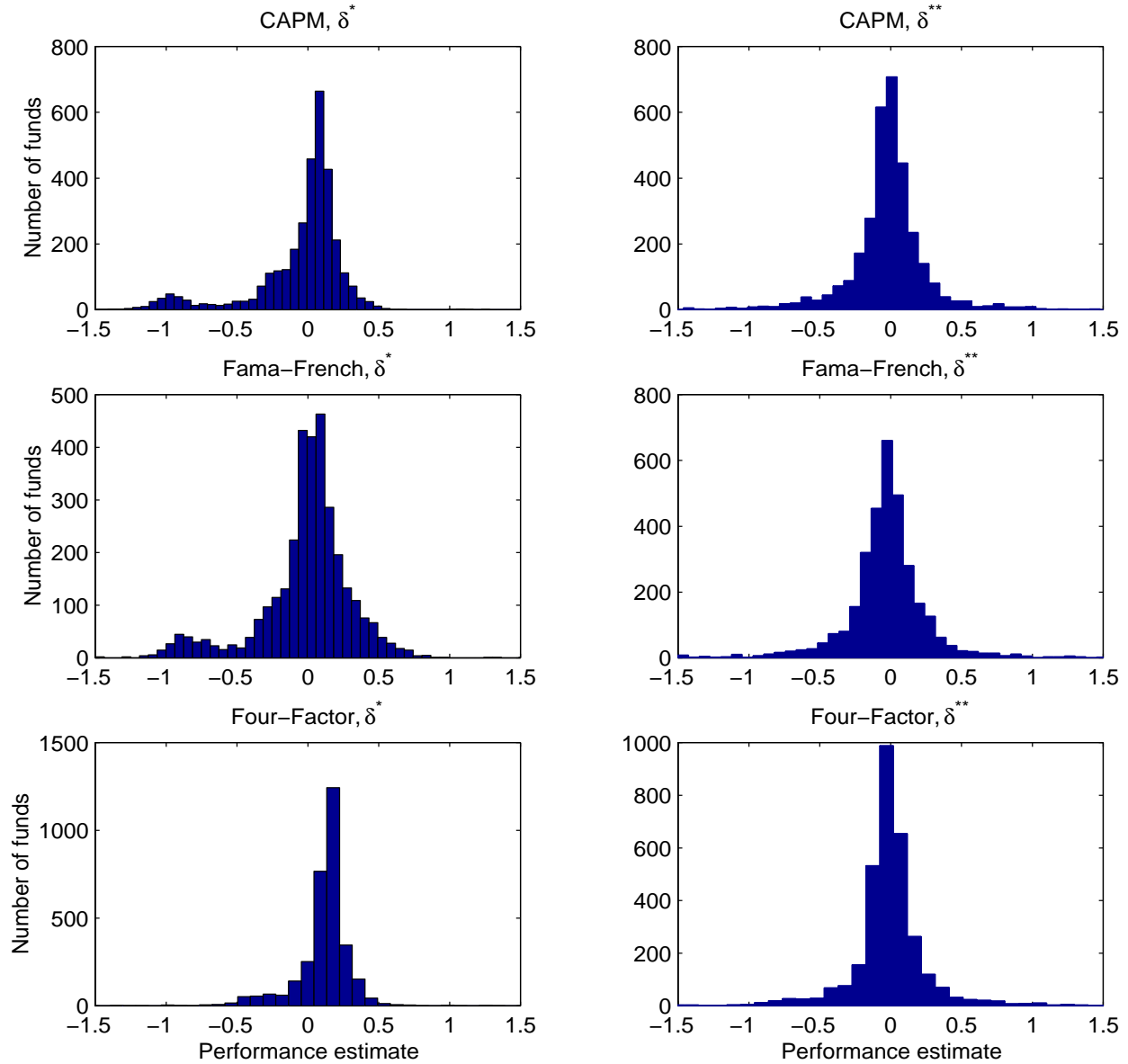


Figure 2. Empirical distribution of our performance measures across funds.

The figure plots the histograms of our performance measures for all funds with valid data at the end of our sample. In the top two plots, the reference measure is the CAPM alpha, in the middle plots, it is the Fama-French alpha, and in the bottom plots, the four-factor alpha. The left panels plot the levels measures $\hat{\delta}^*$, and the right panels plot the changes measures $\hat{\delta}^{**}$. All measures are expressed in percent per month.

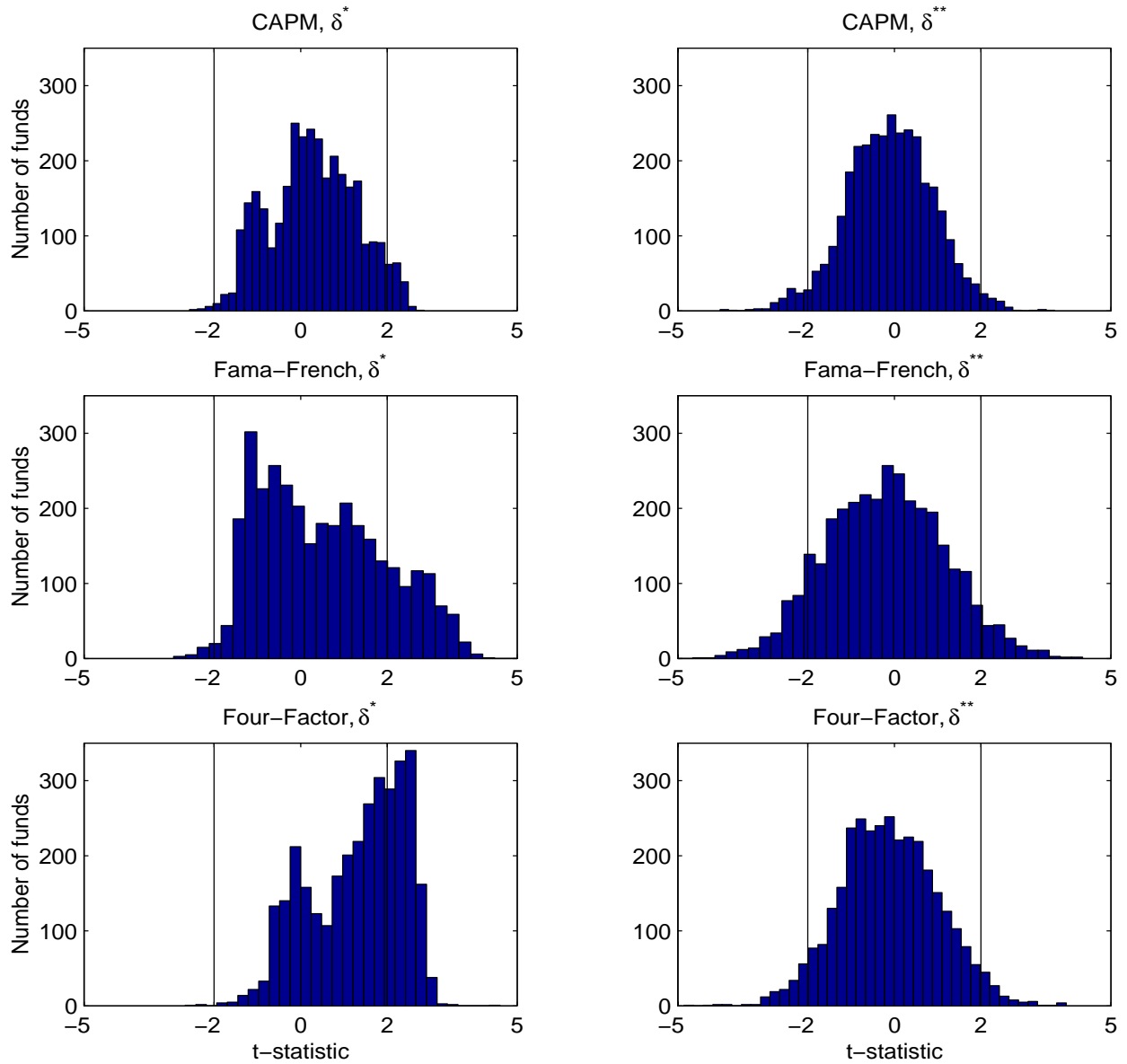


Figure 3. Empirical distribution of the t -statistics of our performance measures across funds.

The figure plots the histograms of the t -statistics of our performance measures for all funds with valid data at the end of our sample. In the top two plots, the reference measure is the CAPM alpha, in the middle plots, it is the Fama-French alpha, and in the bottom plots, the four-factor alpha. The left panels plot the t -statistics of the levels measures $\hat{\delta}^*$, and the right panels plot the t -statistics of the changes measures $\hat{\delta}^{**}$.

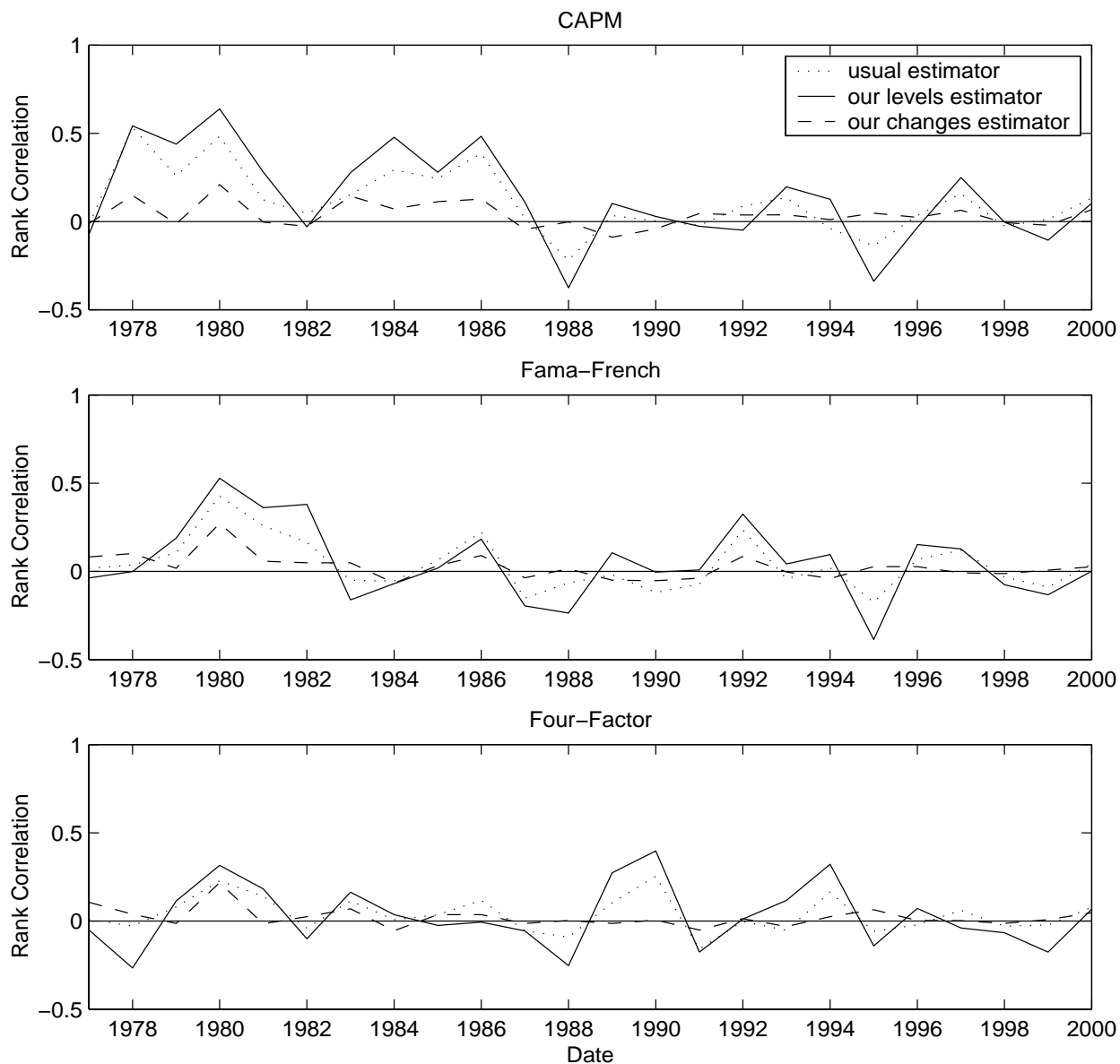


Figure 4. Year-by-year performance persistence.

The figure plots the rank correlations across funds between performance measures estimated at the end of the year given on the horizontal axis and one year earlier. No overlapping return data is used. In each panel, the solid line corresponds to our levels measure $\hat{\delta}^*$, the dashed line to our changes measure $\hat{\delta}^{**}$, and the dotted line to the usual measure $\hat{\delta}$. In the top panel, the reference measure is the CAPM alpha, in the middle panel, it is the Fama-French alpha, and in the bottom panel, the four-factor alpha.

Appendix

Computing $\Omega = \text{Cov}(\hat{\delta}, \hat{\delta}')$ for funds with misaligned return histories.

Let $\mathcal{S} = \{1, \dots, T\}$ denote the set of dates in the whole sample period in which fund returns may be available. Suppose that returns on fund 1 are available in the subset \mathcal{S}_1 of the whole sample period, $\mathcal{S}_1 \subset \mathcal{S}$. These returns are stacked in the vector R_1 , whose dimension is $N_1 \times 1$, where N_1 is the number of observations for fund 1. Analogously, suppose that returns on fund 2 are available in $\mathcal{S}_2 \subset \mathcal{S}$, and they are stacked in the $N_2 \times 1$ vector R_2 . Also let R_B denote the vector of returns on K benchmark portfolios, available for the whole period \mathcal{S} . All returns are in excess of the risk-free rate. Define

$$R_{1,t} = \delta_1 + R_{B,t}\beta_1 + \epsilon_{1,t}, \quad t \in \mathcal{S}_1 \quad (\text{A1})$$

$$R_{2,t} = \delta_2 + R_{B,t}\beta_2 + \epsilon_{2,t}, \quad t \in \mathcal{S}_2 \quad (\text{A2})$$

The estimated intercepts $\hat{\delta}_1$ and $\hat{\delta}_2$ are obtained by running separate OLS regressions, one using the data from \mathcal{S}_1 and the other using the data from \mathcal{S}_2 .¹⁸ The standard errors of $\hat{\delta}_1$ and $\hat{\delta}_2$ are computed accordingly. To calculate Ω , we also need $\text{Cov}(\hat{\delta}_1, \hat{\delta}_2)$ for each pair of funds. Rewrite equations (A1) and (A2) as

$$R_1 = X_1\theta_1 + \epsilon_1 \quad (\text{A3})$$

$$R_2 = X_2\theta_2 + \epsilon_2, \quad (\text{A4})$$

where $\theta_j = (\delta_j \ \beta_j')$, X_j is the subset of $X = [\iota_T \ R_B]$ corresponding to \mathcal{S}_j (i.e. we take only the rows that correspond to the dates in \mathcal{S}_j), $j = 1, 2$, and ι_T is a T -vector of ones. Then

$$\hat{\theta}_1 = (X_1'X_1)^{-1}X_1'R_1 = (X_1'X_1)^{-1}X_1'(X_1\theta_1 + \epsilon_1) = \theta_1 + (X_1'X_1)^{-1}X_1'\epsilon_1$$

$$\hat{\theta}_2 = (X_2'X_2)^{-1}X_2'R_2 = (X_2'X_2)^{-1}X_2'(X_2\theta_2 + \epsilon_2) = \theta_2 + (X_2'X_2)^{-1}X_2'\epsilon_2$$

Hence,

$$\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = \text{E}[(\hat{\theta}_1 - \theta_1)(\hat{\theta}_2 - \theta_2)'] = (X_1'X_1)^{-1}X_1'\text{E}(\epsilon_1\epsilon_2')X_2(X_2'X_2)^{-1} \quad (\text{A5})$$

¹⁸Using the techniques of Pástor and Stambaugh (2002), the intercepts of short-history funds can be estimated more precisely by incorporating the returns on longer-history funds. The most efficient use of all information involves regressing fund returns on the returns of all longer-history funds and the benchmarks. To ensure feasibility, we would have to choose a subset of longer-history funds, and the estimates would depend on that subset, which seems undesirable. (Pástor and Stambaugh use a small set of carefully-selected longer-history passive assets instead of longer-history funds.) We opt for the simplicity of estimating the regression intercepts fund by fund.

Note that ϵ_1 is $N_1 \times 1$ and ϵ_2 is $N_2 \times 1$, so that $E(\epsilon_1 \epsilon_2')$ is $N_1 \times N_2$. Let σ_{12} denote the contemporaneous covariance between ϵ_1 and ϵ_2 . Then $E(\epsilon_1 \epsilon_2')$ is a matrix whose (i, j) element is σ_{12} if $\mathcal{S}_1(i) = \mathcal{S}_2(j)$ and zero otherwise, since the epsilons are assumed to be uncorrelated over time. Let \mathcal{O} denote the overlap of the funds' sample periods, $\mathcal{O} = \mathcal{S}_1 \cap \mathcal{S}_2$, let $X_{\mathcal{O}}$ denote the row subset of X corresponding to \mathcal{O} , let $N_{\mathcal{O}}$ denote the number of elements in \mathcal{O} , and let $I_{N_{\mathcal{O}}}$ denote the identity matrix of dimension $N_{\mathcal{O}}$. Equation (A5) can be rewritten as

$$\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = (X_1' X_1)^{-1} X_1' (\sigma_{12} I_{N_{\mathcal{O}}}) X_{\mathcal{O}} (X_2' X_2)^{-1} \quad (\text{A6})$$

$$= \sigma_{12} (X_1' X_1)^{-1} (X_{\mathcal{O}}' X_{\mathcal{O}}) (X_2' X_2)^{-1}. \quad (\text{A7})$$

Our estimate of $\text{Cov}(\hat{\delta}_1, \hat{\delta}_2)$ is given by the $(1, 1)$ element of (A7). (As an example, if the history of fund 2 is subsumed by the history of fund 1, $\mathcal{S}_2 \subset \mathcal{S}_1$, so that $\mathcal{O} = \mathcal{S}_2$, then $\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = \sigma_{12} (X_1' X_1)^{-1}$.) To estimate σ_{12} , we run the regressions (A1) and (A2) on the overlapping data \mathcal{O} and take the sample covariance of the resulting residuals.¹⁹

Proof of the statement immediately following equation (26).

$$\begin{aligned} \sum_{j=1}^M c_{m,j} &= \sum_{j=1}^M \sum_{n=1}^N \left[x_{m,n}^+ y_{j,n}^+ \mathbf{1}_{\{n \in \mathcal{N}_m^+\}} \mathbf{1}_{\{j \in \mathcal{M}_n^+\}} - x_{m,n}^+ y_{j,n}^- \mathbf{1}_{\{n \in \mathcal{N}_m^+\}} \mathbf{1}_{\{j \in \mathcal{M}_n^-\}} \cdots \right. \\ &\quad \left. - x_{m,n}^- y_{j,n}^+ \mathbf{1}_{\{n \in \mathcal{N}_m^-\}} \mathbf{1}_{\{j \in \mathcal{M}_n^+\}} + x_{m,n}^- y_{j,n}^- \mathbf{1}_{\{n \in \mathcal{N}_m^-\}} \mathbf{1}_{\{j \in \mathcal{M}_n^-\}} \right] \\ &= \sum_{n=1}^N x_{m,n}^+ \mathbf{1}_{\{n \in \mathcal{N}_m^+\}} \sum_{j=1}^M y_{j,n}^+ \mathbf{1}_{\{j \in \mathcal{M}_n^+\}} - \sum_{n=1}^N x_{m,n}^+ \mathbf{1}_{\{n \in \mathcal{N}_m^+\}} \sum_{j=1}^M y_{j,n}^- \mathbf{1}_{\{j \in \mathcal{M}_n^-\}} \cdots \\ &\quad - \sum_{n=1}^N x_{m,n}^- \mathbf{1}_{\{n \in \mathcal{N}_m^-\}} \sum_{j=1}^M y_{j,n}^+ \mathbf{1}_{\{j \in \mathcal{M}_n^+\}} + \sum_{n=1}^N x_{m,n}^- \mathbf{1}_{\{n \in \mathcal{N}_m^-\}} \sum_{j=1}^M y_{j,n}^- \mathbf{1}_{\{j \in \mathcal{M}_n^-\}} \\ &= \sum_{n=1}^N x_{m,n}^+ \mathbf{1}_{\{n \in \mathcal{N}_m^+\}} - \sum_{n=1}^N x_{m,n}^+ \mathbf{1}_{\{n \in \mathcal{N}_m^+\}} - \sum_{n=1}^N x_{m,n}^- \mathbf{1}_{\{n \in \mathcal{N}_m^-\}} + \sum_{n=1}^N x_{m,n}^- \mathbf{1}_{\{n \in \mathcal{N}_m^-\}} \\ &= 0. \end{aligned} \quad (\text{A8})$$

¹⁹There is a more sophisticated approach to estimating σ_{12} when $\mathcal{S}_2 \subset \mathcal{S}_1$. The results in Stambaugh (1997) can be used to estimate σ_{12} using also the data in \mathcal{S}_1 that is not in \mathcal{S}_2 . However, since $\mathcal{S}_2 \subset \mathcal{S}_1$ is unlikely to happen for all pairs of funds and since we want the same procedure for all pairs, we simply take the estimate of σ_{12} from the overlapping data.

Inference and investment of our simulated investors.

Here we calculate the expected return and the covariance matrix of returns as perceived by the simulated mean-variance investors considered in Section II. The subscript m , which denotes investor m , is dropped throughout for convenience.

First note that with a diffuse prior on α_n , the Bayes rule implies that

$$\alpha_n | s_n, \gamma = \begin{cases} s_n & \text{with probability } \gamma \\ u_n \sim N(0, \sigma_\alpha^2) & \text{with probability } 1 - \gamma, \end{cases} \quad (\text{A9})$$

so that

$$\begin{aligned} \mathbb{E}(\alpha_n | s_n, \gamma) &= \gamma s_n \\ \text{Var}(\alpha_n | s_n, \gamma) &= \mathbb{E}(\alpha_n^2 | s_n, \gamma) - [\mathbb{E}(\alpha_n | s_n, \gamma)]^2 = (\gamma - \gamma^2)s_n^2 + (1 - \gamma)\sigma_\alpha^2. \end{aligned}$$

Let $\mathbb{E}(\gamma)$ denote the investor's perception of his own skill, and let $\text{Var}(\gamma)$ denote the variance associated with this perception. Using the law of iterated expectations, the expected return on stock n after observing the signals is equal to

$$\mathbb{E}(r_n | S) = \mathbb{E}[\mathbb{E}(\alpha_n | s_n, \gamma)] = \mathbb{E}(\gamma)s_n. \quad (\text{A10})$$

Using the variance decomposition rule, the perceived variance of stock n 's return is

$$\begin{aligned} \text{Var}(r_n | S) &= \sigma_e^2 + \mathbb{E}[\text{Var}(\alpha_n | s_n, \gamma)] + \text{Var}[\mathbb{E}(\alpha_n | s_n, \gamma)] \\ &= \sigma_e^2 + \sigma_\alpha^2 + \mathbb{E}(\gamma)(s_n^2 - \sigma_\alpha^2) - s_n^2[\mathbb{E}(\gamma)]^2. \end{aligned} \quad (\text{A11})$$

The perceived covariance between returns on stocks i and j is

$$\begin{aligned} \text{Cov}(r_i, r_j | S) &= \mathbb{E}[\text{Cov}(\alpha_i, \alpha_j | S, \gamma)] + \text{Cov}[\mathbb{E}(\alpha_i | S, \gamma), \mathbb{E}(\alpha_j | S, \gamma)] \\ &= \mathbb{E}[\mathbb{E}((\alpha_i - \gamma s_i)(\alpha_j - \gamma s_j) | S, \gamma)] + \text{Cov}[\gamma s_i, \gamma s_j] \\ &= \mathbb{E}[\gamma^2 s_i s_j - \gamma^2 s_i s_j - \gamma^2 s_i s_j + \gamma^2 s_i s_j] + s_i s_j \text{Var}(\gamma) \\ &= s_i s_j \text{Var}(\gamma). \end{aligned} \quad (\text{A12})$$

For simplicity, we assume that each manager knows his own γ , so that $\mathbb{E}(\gamma) = \gamma$ and $\text{Var}(\gamma) = 0$. The expression for the optimal weights in equation (35) follows immediately.

References

- Baks, Klaas P., 2002, On the performance of mutual fund managers, Working paper, Emory University.
- Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, *Journal of Finance* 56, 45–85.
- Berk, Jonathan B., and Richard C. Green, Mutual fund flows and performance in rational markets, Working paper, University of California, Berkeley.
- Bollen, Nicolas P. B., and Jeffrey A. Busse, 2002, “Short-term persistence in mutual fund performance”, Working paper, Vanderbilt University.
- Brown, Stephen J., and William N. Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679–698.
- Busse, Jeffrey A., and Paul J. Irvine, 2002, “Bayesian alphas and mutual fund persistence”, Working paper, Emory University.
- Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.
- Chen, Hsiu-Lang, Narasimhan Jegadeesh, and Russ Wermers, 2000, The value of active mutual fund management: An examination of the stockholdings and trades of fund managers, *Journal of Financial and Quantitative Analysis* 35, 343–368.
- Cohen, Randolph B., Paul A. Gompers, and Tuomo Vuolteenaho, 2002, Who underreacts to cash-flow news? Evidence from trading between individuals and institutions, *Journal of Financial Economics*, forthcoming.
- Daniel, Kent D., Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.
- Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133–157.
- Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Ferson, Wayne E., and Kenneth Khang, 2002, Conditional performance measurement using portfolio weights: Evidence for pension funds, *Journal of Financial Economics* 65, 249–282.
- Goetzmann, William N., and Roger G. Ibbotson, 1994, Do winners repeat? Patterns in mutual fund performance, *Journal of Portfolio Management* 20, 9–18.
- Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* 62, 394–415.

- Grinblatt, Mark, and Sheridan Titman, 1992, The persistence of mutual fund performance, *Journal of Finance* 47, 1977–1984.
- Grinblatt, Mark, and Sheridan Titman, 1993, Performance measurement without benchmarks: An examination of mutual fund returns, *Journal of Business* 66, 47–68.
- Grinblatt, Mark, Sheridan Titman, and Russ Wermers, 1995, Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, *American Economic Review* 85, 1088–1105.
- Hendricks, Darryl, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: The persistence of performance 1974-1988, *Journal of Finance* 48, 93–130.
- Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2002, “Thy neighbor’s portfolio: Word-of-mouth effects in the holdings and trades of money managers”, Working paper, Stanford University.
- Jensen, Michael, 1968, The performance of mutual funds in the period 1945–1964, *Journal of Finance* 23, 389–416.
- Jones, Christopher S., and Jay Shanken, 2002, Mutual fund performance with learning across funds, Working paper, University of Southern California.
- Kothari, S.P., and Jerold B. Warner, 2000, Evaluating mutual fund performance, *Journal of Finance* 56, 1985-2010.
- Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1992, The impact of institutional trading on stock prices, *Journal of Financial Economics* 32, 23-44.
- Malkiel, Burton, 1995, Returns from investing in mutual funds 1971 to 1991, *Journal of Finance* 50, 549–572.
- Pástor, Ľuboš, and Robert Stambaugh, 2002, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63, 315–349.
- Sharpe, William F., 1966, Mutual fund performance, *Journal of Business* 39, 119–138.
- Shumway, Tyler, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327–340.
- Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.
- Stambaugh, Robert F., 2002, Inference about survivors, Working paper, Wharton.
- Wermers, Russ, 1997, Momentum investment strategies of mutual funds, performance persistence, and survivorship bias, Working paper, University of Maryland.
- Wermers, Russ, 1999, Mutual fund herding and the impact on stock prices, *Journal of Finance* 54, 581-622.
- Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses, *Journal of Finance* 55, 1655–1695.