Journal of Empirical Finance 10 (2003) 399–425
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

A nonparametric test of market timing

Wei Jiang*

Finance and Economics Division, Columbia Business School, 3022 Broadway, New York, NY 10027, USA

Abstract

In this paper, we propose a nonparametric test for market timing ability and apply the analysis to a large sample of mutual funds that have different benchmark indices. The test statistic is formed to proxy the probability that a manager loads on more market risk when the market return is relatively high. The test (i) only requires the ex post returns of funds and their benchmark portfolios; (ii) separates the quality of timing information a money manager possesses from the aggressiveness with which she reacts to such information; and (iii) is robust to different information and incentive structures, as well as to underlying distributions. Overall, we do not find superior timing ability among actively managed domestic equity funds for the period of 1980–1999. Further, it is difficult to predict funds' timing performance from their observable characteristics.

JEL classification: G1; C1
Keywords: Mutual funds; Market timing; Nonparametric test; *U*-statistics

1. Introduction

Based on the theory of market efficiency with costly information, there has been ample research work on measuring professional money managers' performance. The emphasis has been on one of the two basic abilities: securities selectivity and market timing. The former tests whether a fund manager's portfolio outperforms the benchmark portfolio in risk-adjusted terms (Jensen, 1972; Gruber, 1996; Ferson and Schadt, 1996; Kothari and Warner, in press). The latter tests whether a fund manager can out-guess the market by moving in and out of the market (Treynor and Mazuy, 1966; Henriksson and Merton, 1981; Admati et al., 1986; Bollen and Busse, 2001).

* Tel.: +1-212-854-9679; fax: +1-212-316-9180.
E-mail address: wj2006@columbia.edu (W. Jiang).

0927-5398/03/$ - see front matter
doi:10.1016/S0927-5398(02)00065-8

Measures of market timing have fallen into one of the two categories: portfolio- and return-based methods. The former tests whether money managers successfully allocate monies among different classes of assets (e.g., equity versus cash) to capitalize on market ascendancy and/or to avoid downturns. If we could observe the portfolio composition of mutual funds at the same frequency as we observe the returns, we could infer funds' market timing by testing whether the portfolio holdings anticipate market moves. Graham and Harvey (1996) empirically test market timing using investment newsletters' asset allocation recommendations.

Holdings, however, are often not available (especially in academic studies), which limits the market timing analysis to the returns of funds and benchmark portfolios only. The return-based method, on the other hand, only requires data on the ex post returns of funds and the relevant market indices. The two most popular methods along this line are those proposed by Treynor and Mazuy (1966) (henceforth "TM") and Henriksson and Merton (1981) (henceforth "HM").

Most of the work on mutual fund performance measurement extends the CAPM or multi-factor analysis of securities and portfolios to mutual funds. There has been controversy over using such a metric to evaluate mutual fund performance. The static $\alpha-\beta$ analysis misses the diversified and dynamic aspects of managed portfolios (Admati et al., 1986; Ferson and Schadt, 1996; Becker et al., 1999; Ferson and Khang, 2001). A fund manager may vary her portfolio's exposure to the market or other risk factors, or alter the fund's correlation to the benchmark index in response to the incentive she faces (Chevalier and Ellison, 1997). Consequently, the systematic part of the fund's risk can be mis-estimated when its manager is trying to time the market, and existing measures may incorrectly attribute performance to funds, or fail to attribute superior returns to an informed manager (Grinblatt and Titman, 1989). To address these issues, there has been a great deal of study on capturing the effect of conditioning information on timing performance measures (Ferson and Schadt, 1996; Becker et al., 1999; Ferson and Khang, 2001), controlling for spurious timing arising from not holding the benchmark (Jagannathan and Korazjczyk, 1986; Breen et al., 1986), decomposing abnormal performance into selectivity and timing (Admati et al., 1986; Grinblatt and Titman, 1989), and minimizing the loss of test power due to sampling frequencies (Goetzmann et al., 2000; Bollen and Busse, 2001).

In this paper, we develop an independent test to measure the market timing ability of portfolio managers without resorting to the estimation of α's or β's. The test is based on the simple idea that a successful market timer's fund rises significantly when the market rises and falls slightly when the market drops. The nonparametric test has the following properties. First, it is easy to implement because it only requires the ex post returns of funds and their benchmark portfolios. Second, the test statistic is not affected by the manager's risk aversion because it separates the quality of timing information a fund manager possesses from the aggressiveness of the reaction to such information. Third, the test is more robust to different information and incentive structures, as well as to timing frequencies and underlying distributions, than existing timing measures. Finally, the method developed in this paper is readily applicable to analyzing the market timing ability of financial advisors or newsletters (Graham and Harvey, 1996), or the timing behavior of individual investors (Odean, 1998; Barber and Odean, 2000).

The rest of the paper is organized as follows: Section 2 presents the nonparametric statistic of market timing and compares it with the TM and HM methods. Section 3 applies the method to a data set of mutual funds with different benchmark indices. Section 4 concludes.

2. Model

2.1. Market timing test statistics

We assume that a money manager's timing information is independent of her information about individual securities. This is a fairly standard assumption in the performance measurement literature (e.g., see Admati et al., 1986; Grinblatt and Titman, 1989).[1] With independent selectivity and timing, we have the following market model of fund returns (all returns are expressed in excess of the risk-free rate):

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

where i is the subscript for individual funds throughout this paper. $\beta_{i,t}$ is a random variable adapted to the information available to the manager at time t and r_m represents the return of the relevant market (which can be a subset of the total market) in which the mutual fund invests. It is the benchmark portfolio return against which the fund is evaluated. In the simplest case, a market timer decides on β_t at date t and invests β_t percent in the market portfolio and the rest in bonds until date $t+1$. Eq. (1) represents the return process from such a timing strategy.

For a triplet $\{r_{m,t_1}, r_{m,t_2}, r_{m,t_3}\}$ sampled from any three periods such that $r_{m,t_1} < r_{m,t_2} < r_{m,t_3}$, an informed manager should, on average, maintain a higher average β in the $\lfloor r_{m,t_2}, r_{m,t_3} \rfloor$ range than in the $\lfloor r_{m,t_1}, r_{m,t_2} \rfloor$ range. The β estimates for both ranges (given two observations for each range) are $(r_{i,t_2} - r_{i,t_1})/(r_{m,t_2} - r_{m,t_1})$ and $(r_{i,t_3} - r_{i,t_2})/(r_{m,t_3} - r_{m,t_2})$, respectively. Accordingly, we propose using the probability

$$\theta = 2\Pr\left(\frac{r_{i,t_3} - r_{i,t_2}}{r_{m,t_3} - r_{m,t_2}} > \frac{r_{i,t_2} - r_{i,t_1}}{r_{m,t_2} - r_{m,t_1}}\right) - 1, \tag{2}$$

as a statistic of market timing ability. We motivate this market timing measure as follows.

A manager's timing ability is determined by the relevance and accuracy of her information. Let $\hat{r}_{m,t+1} = E(r_{m,t+1} \mid I_t)$ be the manager's prediction about the next-period market return based on I_t, her information set (both public and private) at time t. If I_t is not informative at all, then the conditional distribution equals the unconditional one, that is, $f(r_{m,t+1} | \hat{r}_{m,t+1}) = f(\hat{r}_{m,t+1})$, where $f(\cdot)$ stands for the probability density function. In this case, the conditional forecast equals the unconditional one and the manager would not be able to tell when the market will enjoy relatively high returns. More specifically, for two

[1] Correlated timing and selectivity information would in general cause technical difficulties in separating abnormal performance due to timing from that due to selectivity. For a detailed discussion, see Grinblatt and Titman (1989).

periods, $t_1 \neq t_2$, the following parameter takes the value of zero in the absence of timing information[2]:

$$\begin{aligned} v &= \Pr(\hat{r}_{m,t_1+1} > \hat{r}_{m,t_2+1} \mid r_{m,t_1+1} > r_{m,t_2+1}) - \Pr(\hat{r}_{m,t_1+1} < \hat{r}_{m,t_2+1} \mid r_{m,t_1+1} > r_{m,t_2+1}) \\ &= 2\Pr(\hat{r}_{m,t_1+1} > \hat{r}_{m,t_2+1} \mid r_{m,t_1+1} > r_{m,t_2+1}) - 1. \end{aligned} \tag{3}$$

At the other extreme, if the forecast is always perfect, that is, $\hat{r}_{m,t+1} \equiv r_{m,t+1}$, then v attains its upper bound of one. Symmetrically, $v = -1$ represents perfectly perverse market timing. Therefore, the value of $v \in [-1,1]$ indicates the fund manager's market timing ability: the more accurate the information I_t the higher the value of v. The next step is to find a relationship between the manager's forecast ($\hat{r}_{m,t+1}$) and her action (β_t) so that θ defined in Eq. (2) is a valid proxy of v.

Suppose the manager receives a favorable signal that leads to a high $\hat{r}_{m,t+1}$. How much market exposure (β_t) the manager would like to take apparently depends on two factors: the precision of the forecast and the aggressiveness with which she uses her own information. The first part concerns natural ability, while the latter can be affected by the manager's risk aversion. Grinblatt and Titman (1989) show that an investor who has independent timing and selectivity information and non-increasing absolute risk aversion[3] would increase β_t in Eq. (1) as information about the market becomes more favorable, or $\frac{\partial \beta_t}{\partial \hat{r}_{m,t+1}} > 0$. Combining $\frac{\partial \beta_t}{\partial \hat{r}_{m,t+1}} > 0$ with Eq. (3), we see that the following probability is greater than zero if and only if the manager possesses superior timing information:

$$2\Pr(\beta_{t_1} > \beta_{t_2} \mid r_{m,t_1+1} > r_{m,t_2+1}) - 1. \tag{4}$$

From the analysis above, therefore, superior timing ability $v>0$ (defined in Eq. (2)) translates into $\theta>0$ (defined in Eq. (2)) if a manager loads on more market risk when signals about future market returns are more favorable. Eq. (2) is testable because the sample analogue of θ can be formed. Under the null hypothesis of no timing ability, the β has no correlation with the market return, in which case the statistic θ assumes the neutral value of zero. Intuitively, an uninformed manager would move the market exposure of her portfolio in the right direction as often as she would do in the wrong direction. Note that a triplet $\{r_{i,t_1}, r_{i,t_2}, r_{i,t_3}\}$ is convex vis-à-vis the market return if and only if $(r_{i,t_3} - r_{i,t_2})/(r_{m,t_3} - r_{m,t_2}) > (r_{i,t_2} - r_{i,t_1})/(r_{m,t_2} - r_{m,t_1})$. Therefore, θ measures the *probability* that the fund returns bear a convex relation with the market returns in excess of that of a concave relation.

[2] The HM method tests whether the probability $\Pr(\hat{r}_{m,t+1} > 0 \mid r_{m,t+1} > 0) + \Pr(\hat{r}_{m,t+1} < 0 \mid r_{m,t+1} < 0)$ is greater than one. When the HM model is the correct specification, our measure picks up the manager's timing ability among a subset of triplets where at least two observations of market returns are of opposite signs. In general, our measure allows the manager to make finer forecasts and uses more information in the return data by looking at all triplets $\{r_{m,t_1+1}, r_{m,t_2+1}, r_{m,t_3+1}\}$ for $t_1 \neq t_2 \neq t_3$.

[3] Non-increasing absolute risk aversion requires that the investor's risk aversion measured by $-\frac{u''(w)}{u'(w)}$ be non-increasing in the wealth level w. Commonly used utility functions, such as the exponential, power, and log utilities, all meet this criterion.

The sample analogue to θ becomes a natural candidate as a statistic. It is a U-statistic with kernel of order three:

$$\hat{\theta}_n = \binom{n}{3}^{-1} \sum_{r_{m,t_1} < r_{m,t_2} < r_{m,t_3}} \mathrm{sign}\left(\frac{r_{i,t_3} - r_{i,t_2}}{r_{m,t_3} - r_{m,t_2}} > \frac{r_{i,t_2} - r_{i,t_1}}{r_{m,t_2} - r_{m,t_1}} \right), \tag{5}$$

where n is the sample size and sign(·) is the sign function that assumes value 1 (− 1) if the argument is positive (negative) and equals zero if the argument is zero. By the property of U-statistics, $\hat{\theta}_n$ is a $\sqrt{n}$-consistent and asymptotically normal estimator for θ (Serfling, 1980; Abrevaya and Jiang, 2001). That is, $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2_{\hat{\theta}_n})$ when $n \to \infty$. Further $\hat{\theta}_n$, as defined in Eq. (5), is the least variance estimator among all unbiased estimators for the population coefficient θ.

Abrevaya and Jiang (2001) provide the asymptotic distribution of the $\hat{\theta}_n$ statistic. Let $z_{t_j} \equiv (r_{t_j}, r_{m,t_j})$, j={1, 2, 3}, and denote the kernel function of $\hat{\theta}_n$ by

$$h(z_{t_1}, z_{t_2}, z_{t_3}) = \mathrm{sign}\left(\frac{r_{i,t_3} - r_{i,t_2}}{r_{m,t_3} - r_{m,t_2}} > \frac{r_{t_2} - r_{i,t_1}}{r_{m,t_2} - r_{m,t_1}} \mid r_{m,t_1} < r_{m,t_2} < r_{m,t_3} \right).$$

A consistent estimator of the standard error of $\hat{\theta}_n$ is derived in Abrevaya and Jiang (2001):

$$\hat{\sigma}^2_{\hat{\theta}_n} = \frac{9}{n} \sum_{t_1=1}^{n} \left(\binom{n}{2}^{-1} \sum_{t_2, t_3} h(z_{t_1}, z_{t_2}, z_{t_3}) - \hat{\theta}_n \right)^2. \tag{6}$$

Simulation results in Abrevaya and Jiang (2001) show that the size of the test is very accurate[4] if we use the bootstrap method in standard error estimation for sample sizes below 50 and use the asymptotic formula for larger sample sizes.

2.2. Properties

The new market timing measure (θ) has a ready interpretation as the probability that a fund manager takes relatively more systematic risk in a higher return period than in a low return one. Since the seminal work of Treynor and Mazuy (1966) and Henriksson and Merton (1981), there has been much work extending these measures in order to relax their restrictive behavioral and distribution assumptions while retaining their intuitive appeal, ease of implementation, and minimal data requirements.[5] In this subsection, we discuss the

[4] Using 1000 simulations, rejection rates at 5% significance level are between 4.5% and 5.5% for all error specifications.

[5] Goetzmann et al. (2000) had an excellent review of the research that addresses the limitation of the TM and HM timing measures.

contribution of the nonparametric timing measure on these grounds and point out its limitations. The fund subscript i will henceforth be omitted where there is no confusion.

2.2.1. Information structure and behavioral assumptions

The nonparametric measure allows a more flexible specification of a fund manager's response to information. We require β_t to be a non-decreasing function of $\hat{r}_{m,t+1}$, that is, the manager sets a higher β for the fund when her forecast of the next-period market return is more favorable. Grinblatt and Titman (1989) show that sufficient conditions for this to hold are *i.i.d.* random noise in market returns, independent selectivity and timing information, and non-increasing absolute risk aversion. This requirement is less stringent than those of the TM and HM measures, which require linear or binary response function by the manager. The *i.i.d.* assumption, however, rules out heteroscedasticity in returns and hence volatility timing by money managers. We will relax this assumption and discuss the possible impact of volatility timing in a later section.

In general, a fund manager's reaction to information depends on her risk aversion (which could be affected by the incentive she faces) as well as her natural ability. The functional form of such a response is difficult to specify without being somewhat arbitrary. For example, the TM measure uses the following quadratic regression of a fund's returns:

$$r_{t+1} = \alpha + \beta r_{m,t+1} + \gamma[r_{m,t+1}]^2 + \varepsilon_{t+1}, \tag{7}$$

where superior timing shows up in a positive coefficient γ_i. As analyzed in Admati et al. (1986), the return process of Eq. (7) comes out of a linear response by the fund manager in the form of:

$$\beta_t = \bar{\beta} + \lambda[\hat{r}_{m,t+1} - E(r_m)]. \tag{8}$$

The linear response function is consistent with the manager's acting as if she were maximizing the expected utility of a CARA preference. However, such an assumption is questionable if the fund manager maximizes the utility related to her own payoff under the incentive she faces instead of the fund's total return. The deviation from maximizing a CARA preference is large when there is non-linearity in the incentive, explicitly or implicitly, in the forms of benchmark evaluation (Admati and Pfleiderer, 1997), option compensation (Carpenter, 2000), or non-linear flow-to-performance responses by fund investors (Chevalier and Ellison, 1997).

The HM measure, on the other hand, assumes that a manager takes only two β values—a high β when she expects the market return to exceed the risk-free rate and a low β when otherwise. The binary-β strategy results in the following return model:

$$r_{t+1} = \alpha + \beta r_{m,t+1} + \gamma[r_{m,t+1}]^+ + \varepsilon_{t+1}, \tag{9}$$

where $[r_{m,t+1}]^+ = \max(0, r_{m,t+1})$. The coefficient on $[r_{m,t+1}]^+$ represents the value added by effective timing that is equivalent to a call option on the market portfolio where the exercise price equals the risk-free rate. Such a specification, while intuitive, is highly restrictive as well. After all, there is no reason to expect a uniform reaction to information by all fund managers. In comparison, the nonparametric measure offers more flexibility. It only requires the reaction function to be non-decreasing in the manager's forecast of market return.

When a linear reaction function is the correct specification, the nonparametric measure gives the same result as the TM measure. In the TM model, the manager's private signal, y_t, is generated according to

$$y_t = r_{m,t+1} + \eta_t, \tag{10}$$

where η_t is a normal random variable that is independent of $r_{m,t+1}$ and is *i.i.d.* across time. Timing ability is represented by the inverse of the variance of the noise term. For any two η_{t_1} and η_{t_2} from two periods $t_1 \neq t_2$, we can calculate Eq. (3) as follows:

$$\begin{aligned} v &= 2\Pr(\eta_{t_2} - \eta_{t_1} < r_{m,t_1+1} - r_{m,t_2+1} \mid r_{m,t_1+1} > r_{m,t_2+1}] - 1 \\ &= 2\Phi\left(\frac{|r_{m,t_1+1} - r_{m,t_2+1}|}{\sqrt{2}\sigma_\eta}\right) - 1, \end{aligned} \tag{11}$$

where $\Phi(\cdot)$ stands for the cumulative probability function of the standard normal distribution. It is easy to see that v is monotonically increasing in $1/\sigma_\eta$, the precision of the private signal. An infinitely noisy signal ($\sigma_\eta = \infty$) leads to $v=0$ (no timing) and a perfect signal ($\sigma_\eta = 0$) implies $v=1$ (perfect timing). Therefore, the nonparametric measure will identify a good timer who adopts the TM timing strategy.

2.2.2. Ability versus response

A fund manager's market timing performance relies on both the quality of her private information (ability) and the aggressiveness with which the manager reacts to her information (response). This constitutes a dichotomy that is difficult to decompose. Except for special cases, existing performance measures are not able to extract the information-related component of performance. As Grinblatt and Titman (1989) point out, it would be better if performance measures (in addition to detecting abnormal performance) could "also select the more informed of two [managers]". An investor should be more concerned with the quality of the manager's information than with the manager's aggressiveness because the investor can choose the proportion of her wealth invested in the fund in response to the manager's ability.

The TM and HM measures reflect both aspects of market timing. We see that the estimated $\hat{\gamma}_{TM}$ in the TM regression will pick up the coefficient in the linear reaction function (the λ term in Eq. (8)). Hence, more aggressive funds can show up with higher $\hat{\gamma}_{TM}$. The $\hat{\gamma}_{TM}$ coefficient in the HM model is an unbiased estimate for the product $\Delta(\beta_H - \beta_L)$, where Δ is the probability defined in footnote 4, and $\beta_H(\beta_L)$ is the manager's target β when the predicted market excess return is positive (negative). Thus, both ability (the Δ term) and aggressiveness (the $\beta_H - \beta_L$ term) are reflected in the estimated timing. The nonparametric statistic, on the other hand, measures how often a manager correctly ranks a market movement and appropriately acts on it, instead of measuring how aggressively she acts on it. We see that, in the linear response case (as in Eq. (8)), the λ coefficient cancels out in the nonparametric measure because

$$\theta = 2\Pr\left(\lambda\frac{\hat{r}_{m,t_3}r_{m,t_3} - \hat{r}_{m,t_2}r_{m,t_2}}{r_{m,t_3} - r_{m,t_2}} > \lambda\frac{\hat{r}_{m,t_2}r_{m,t_2} - \hat{r}_{m,t_1}r_{m,t_1}}{r_{m,t_2} - r_{m,t_1}} \mid r_{m,t_1} < r_{m,t_2} < r_{m,t_3}\right) - 1$$
$$= 2\Pr\left(\frac{\hat{r}_{m,t_3}r_{m,t_3} - \hat{r}_{m,t_2}r_{m,t_2}}{r_{m,t_3} - r_{m,t_2}} > \frac{\hat{r}_{m,t_2}r_{m,t_2} - \hat{r}_{m,t_1}r_{m,t_1}}{r_{m,t_2} - r_{m,t_1}} \mid r_{m,t_1} < r_{m,t_2} < r_{m,t_3}\right) - 1.$$

Thus, our measure largely reflects the information quality component of performance. Based on this analysis, we also see that there is great complementarity between the nonparametric method and the two other methods. Used together in empirical work, they can offer a more complete picture of the market timing performance of fund managers.

2.2.3. Conditional information

The nonparametric measure can be extended to the context of conditional market timing. The literature on conditional performance evaluation stresses the importance of distinguishing performance that merely reflects publicly available information (as captured by a set of instrumental variables) from performance that can be attributed to better information. The conditional market timing approach (see, e.g., Ferson and Schadt, 1996; Graham and Harvey, 1996; Becker et al., 1999; Ferson and Khang, 2001) assumes that investors can time the market on their own using readily available public information, or that by trading on other accounts they can undo any perverse timing that is predicted from the public information. Under such circumstances, the real contribution of a fund manager would be successful timing on the residual part of market returns that is not predictable from public information.

Let $\tilde{r}_{m,t}$ and $\tilde{r}_{\mathrm{i},t_j}$, $j=1, 2, 3$, be the residuals of market returns and the fund return that cannot be explained by lagged instrumental variables. The following statistic then proxies the probability that a fund manager loads on more market risk when the market return is higher, controlled for public information in both market and fund returns:

$$\tilde{\theta}_n = \binom{n}{3}^{-1} \sum_{\tilde{r}_{m,t_1}<\tilde{r}_{m,t_2}<\tilde{r}_{m,t_3}} \mathrm{sign}\left(\frac{\tilde{r}_{i,t_3} - \tilde{r}_{i,t_2}}{\tilde{r}_{m,t_3} - \tilde{r}_{m,t_2}} > \frac{\tilde{r}_{i,t_2} - \tilde{r}_{i,t_1}}{\tilde{r}_{m,t_2} - \tilde{r}_{m,t_1}}\right). \qquad (12)$$

Theoretically, θ in Eq. (2) and $\tilde{\theta}$ in Eq. (12) can have different magnitudes or even different signs because the probabilities are conditional on different states. That is, a manager who successfully times the unpredicted part of the market return can show up as a mis-timer on the gross market return if we do not control for public information. Both public and private information can be used to enhance portfolio returns, but a truly informed manager should have superior market timing based on information beyond that which is readily available to the public.

2.2.4. Statistical robustness

Breen et al. (1986) point out that heteroscedasticity can significantly affect the conclusions of the HM tests. Jagannathan and Korajczyk (1986) and Goetzmann et al. (2000) demonstrate the bias of the HM measure due to skewness. The asymptotic distribution of the $\hat{\theta}_n$ statistic, on the other hand, is unaffected by heteroscedasticity or skewness. Further, $\hat{\theta}_n$ in Eq. (5) is the least variance estimator among all unbiased estimators of θ in Eq. (2). The simulation results shown in Abrevaya and Jiang (2001) demonstrate that the nonparametric test has accurate size even for small samples and is robust (in terms of both the value of the statistic and its standard error) to outliers, non-normality, and heteroscedasticity that are common in financial data.[6] However, we do require the errors in Eq. (1) to be serially uncorrelated. As we will be using monthly return data for our empirical test, this assumption is not a serious concern. However, the statistic can be biased when applied to high-frequency data.[7]

The nonparametric method also offers a timing measure that has little correlation with the estimation error in the standard selectivity measures. TM or HM type regression models would produce a spurious negative correlation between estimated selectivity and timing because of the negatively correlated sampling errors between the two estimates (Jagannathan and Korajczyk, 1986; Coggin, 1993; Kothari and Warner, in press). Our simulation shows that a significant negative correlation between the two estimated abilities will occur in the TM or HM models (or between the selectivity measure from one model and the timing measure from the other) even when the correlation is non-existent. Coggin et al. (1993) and Goetzmann et al. (2000) have similar results. On the other hand, the correlation between $\hat{\theta}_n$ and the selectivity measures from standard regression models is close to the truth.

2.2.5. Model specification and potential bias

In this section, we discuss three specification issues that can affect the consistency and power of market timing tests: the separability of timing from selectivity; the difference between the frequencies at which data are sampled and at which the manager times the market; the relationship between market timing and volatility timing. The nonparametric measure is more robust to model specifications than the TM and HM measures, though it does not overcome all the biases.

A manager can enhance portfolio returns by selecting securities and by timing the market. Decomposing returns in this fashion, however, is empirically difficult (Admati et al., 1986; Grinblatt and Titman, 1989; Coggin et al., 1993; Kothari and Warner, in press). Our measure relies on two common assumptions to avoid detecting spurious timing because of selectivity issues. The first assumption is that a portfolio manager's information on the selectivity side (movement of individual securities) is independent of her

[6] For example, Bollen and Busse (2001) test the hypothesis that fund returns are normally distributed and reject normality at the 1% level. They also conjecture that the relative skewness of market and fund returns is driven by the crash of 1987 and other smaller crashes in the sample.

[7] When applying the measure to high-frequency data, we would recommend the following modification in forming $\hat{\theta}_n$: use only triplet observations $\{r_{m,t_1+1}, r_{m,t_2+1}, r_{m,t_3+1}\}$ that are at least k periods apart, where k is the lag of possible serial correlation, and rescale the statistics by the number of triplets actually used, denote it m. For any finite k, $m \rightarrow \binom{n}{3}^{-1}$ when $n \rightarrow \infty$.

information on the timing side (market movement). In practice, this requires that each individual security constitutes only a small portion of a diversified portfolio and has a negligible impact on the whole market (the manager does not select "too many" stocks at one time, either); or the fund manager must act on selectivity at a much lower frequency than on market timing (so that the manager keeps roughly constant the composition of her risky portfolio when trying to time the market). The second assumption is that the portfolio does not contain derivatives. Jagannanthan and Korajczk (1986) show that buying call options, for example, can induce spurious timing ability. Kosik and Pontiff (1999) find that 21% of the 679 domestic equity funds in their sample hold derivative securities, but detailed information about their derivative holdings is not available. Our measure, like the TM and HM measures, cannot distinguish market timing from option-related spurious timing.

For most timing measures, biases arise when the econometrician observes return data at a frequency different from the frequency at which the manager times the market. Goetzmann et al. (2000) show that monthly evaluation of daily timers using the HM measure is biased severely downward. At the same time, a major component of timing skill would show up as security-selection skill. Bollen and Busse (2001) show that the results of standard timing tests are sensitive to the frequency of data used. Ferson and Khang (2001) point out that an "interim trading bias" can arise when expected returns are time varying and managers trade between return observation dates. The major source of bias is the mis-specification of the regressor $[r_m]^+$ in the HM equation that should take different values depending on the actual timing frequency rather than uniform frequencies (such as monthly). Goetzmann et al. (2000) suggest replacing the monthly option value $[r_m]^+$ with its accumulated daily option value when daily data of fund returns are not readily available. Simulations show that the nonparametric measure is more robust to the difference between timing frequency and sampling frequency because it does not rely on a regression involving a potentially unknown regressor $[r_m]^+$ measured at the "right" frequency. Ferson and Khang (2001) use conditional portfolio weights to control for interim trading bias as well as for trading on public information. Since our measure does not use portfolio weights, it can potentially be subject to such bias.

The third model specification issue comes from the fact that the manager might be timing market volatilities as well as market returns. Busse (1999) shows that funds attempt to decrease market exposure when market volatility is high. Laplante (2001) shows that observed mutual fund positions are not informative about future market volatility. If volatility and expected return are uncorrelated, then our market timing measure remains consistent in the presence of volatility timing. If the correlation is positive, the market timing measure would underestimate the information quality of a successful volatility timing manager.[8] The opposite is true when the relation is negative. Research on the relationship between the expected return and volatility (see, e.g., Breen et al., 1989; Glosten et al., 1993; Busse, 1999) finds that the relation between return and volatility is weak, both conditionally and unconditionally. If this is the case, the manager's timing on return and volatility likely to be weakly related.

[8] If the manager tries to time the volatility, she may reduce market exposure even when the expected return is high, if high-expected return tends to go with high volatility.

2.3. Simulations

In this subsection, we compare the effectiveness of TM, HM, and nonparametric methods using simulations. We use two sample sizes, 60 and 120 monthly returns. The market returns $r_{m,t}$ are the last 5 or 10 years of the S&P500 monthly returns as of December 1999. The first two unconditional moments are calculated from the data to calibrate $E(r_m)$ and $\sigma(r_m)$. We assume that both $E(r_m)$ and $\sigma(r_m)$ are public knowledge. The funds' return data are generated according to the TM model. It is assumed that the manager receives a private signal in the noise-additive form of Eq. (10). She responds to the signal by making the portfolio β a linear function of the signal as specified in Eq. (8).

The quality of the manager's private information is characterized by σ_η, the standard deviation of the noise term in the signal (η_t). We classify the quality of information into three groups: precise ($\sigma_\eta = 2\sigma(r_m)$), medium ($\sigma_\eta = 4\sigma(r_m)$), and coarse ($\sigma_\eta = 8\sigma(r_m)$). They correspond to the correlation between the signal and the actual market return $\rho(y_t, r_{m,t+1})$ being 0.44, 0.24, and 0.12, respectively.[9] Similarly, we divide the response intensity into three groups: aggressive ($\lambda = 0.20$), medium ($\lambda = 0.10$), and conservative ($\lambda = 0.05$). We apply the three methods, running 1000 simulations for each information-response combination. At medium information quality and a 10-year horizon, the correlation between the nonparametric measure and the TM measure ranges from 0.42 to 0.44; the correlation between the nonparametric measure and the HM measure ranges from 0.57 to 0.60. Both correlations are lower when the manager reacts more aggressively to the signals. The correlation between the TM and HM measures is high and stable at around 0.98. Simulation results of all three measures are posted in Table 1.

From Table 1 (Panel A), we see that the $\hat{\theta}$ estimates are roughly invariant across different response intensities (aggressiveness), but they vary significantly across different levels of information quality (ability). Further, the standard errors of $\hat{\theta}$ are robust across different specifications. The other two methods behave differently. Both $\hat{\gamma}$ coefficients mainly reflect the manager's aggressiveness and display little variation with the quality of information. The intuition behind this separation is that the TM and HM coefficients essentially measure the *expected* convexity in the fund's relation to the market return. That reflects both the *probability* (related to signal quality) and the *magnitude* (related to risk aversion) of the convexity. The nonparametric measure is just the probability, not the magnitude; thus, it largely reflects quality and not risk aversion.[10]

We also examine the spurious correlation between the estimated selectivity ability and timing ability when the correlation is non-existent (for brevity, results are not tabulated). The selectivity measure from the TM model, $\hat{\alpha}_{TM}$, is consistent because the simulation is designed according to the TM specification. We find that the correlation between $\hat{\alpha}_{TM}$ and the timing measures from both the TM and HM models are significantly negative (from -0.65 to -0.75 at medium information quality). This confirms Coggin et al.'s (1993) claim that such a negative correlation is largely an artifact of negatively correlated

[9] Farnsworth et al. (in press) find that the best performing mutual funds (those in the upper 5%) have performance similar to artificial mutual funds that have correlation values ranging from 0.24 to 0.32.

[10] I am grateful to an anonymous referee for suggesting this interpretation.

sampling errors for the two estimates. If we use $\hat{\theta}$ and $\hat{\alpha}_{\mathrm{TM}}$ as proxies for timing and selectivity, the correlation ranges from -0.08 to -0.02 at sample size of 60 and it is indistinguishable from zero when the sample size increases to 120.

Next, we examine the issue of measuring returns at monthly frequency when the true timing frequency is unknown (which can be daily or weekly). We obtain the S&P500 daily data from January 1990 to December 1999 from the Center for Research in Securities Prices (CRSP). We simulate the daily returns of a daily timer according to the TM model using the nine combinations of information quality and manager responsiveness. Then we compound the daily return data into monthly returns and proceed with the estimation using only the monthly data. Results are displayed in Table 1 (Panel B).

Table 1
Comparison of three market timing measures

Panel A: three market timing parameters: what do they measure[a]									
	Nonparametric			Treynor–Mazuy			Henriksson–Merton		
	Precise $\sigma_\eta=2\sigma_{r_m}$	Medium $\sigma_\eta=4\sigma_{r_m}$	Coarse $\sigma_\eta=8\sigma_{r_m}$	Precise $\sigma_\eta=2\sigma_{r_m}$	Medium $\sigma_\eta=4\sigma_{r_m}$	Coarse $\sigma_\eta=8\sigma_{r_m}$	Precise $\sigma_\eta=2\sigma_{r_m}$	Medium $\sigma_\eta=4\sigma_{r_m}$	Coarse $\sigma_\eta=2\sigma_{r_m}$
Sample size = 60									
Aggressive	0.204	0.122	0.060	0.204	0.209	0.208	4.422	4.529	4.495
$\lambda=0.20$	(0.070)	(0.072)	(0.072)	(0.078)	(0.157)	(0.318)	(1.609)	(3.152)	(6.241)
Medium	0.200	0.120	0.062	0.099	0.105	0.109	2.169	2.277	2.335
$\lambda=0.10$	(0.068)	(0.072)	(0.072)	(0.041)	(0.084)	(0.165)	(0.779)	(1.604)	(3.138)
Conservative	0.166	0.106	0.060	0.049	0.047	0.049	1.063	1.073	1.065
$\lambda=0.05$	(0.064)	(0.070)	(0.072)	(0.022)	(0.042)	(0.078)	(0.498)	(0.804)	(1.467)
Sample size = 120									
Aggressive	0.234	0.128	0.066	0.200	0.198	0.203	4.151	4.123	4.189
$\lambda=0.20$	(0.048)	(0.050)	(0.052)	(0.059)	(0.118)	(0.242)	(1.042)	(2.083)	(4.249)
Medium	0.210	0.124	0.062	0.100	0.101	0.093	2.065	2.093	1.966
$\lambda=0.10$	(0.048)	(0.050)	(0.052)	(0.029)	(0.059)	(0.119)	(0.516)	(1.043)	(2.088)
Conservative	0.192	0.110	0.066	0.049	0.049	0.050	1.025	1.019	1.035
$\lambda=0.05$	(0.044)	(0.048)	(0.050)	(0.015)	(0.030)	(0.059)	(0.268)	(0.523)	(1.038)
Panel B: monthly measurement of daily timers[b]									
Aggressive	0.170	0.125	0.078	0.065	0.064	0.065	1.204	1.202	1.201
$\lambda=0.20$	(0.039)	(0.039)	(0.039)	(0.018)	(0.021)	(0.032)	(0.320)	(0.381)	(0.578)
Medium	0.169	0.123	0.073	0.031	0.031	0.031	0.577	0.572	0.569
$\lambda=0.10$	(0.038)	(0.038)	(0.039)	(0.009)	(0.010)	(0.016)	(0.151)	(0.187)	(0.282)
Conservative	0.167	0.124	0.075	0.015	0.015	0.015	0.283	0.283	0.274
$\lambda=0.05$	(0.039)	(0.039)	(0.039)	(0.004)	(0.005)	(0.008)	(0.075)	(0.091)	(0.138)
Panel C: market timing with selectivity[c]									
Aggressive	0.201	0.113	0.064	0.225	0.225	0.229	4.617	4.620	4.736
$\lambda=0.20$	(0.041)	(0.047)	(0.048)	(0.071)	(0.133)	(0.257)	(1.260)	(2.400)	(4.660)
Medium	0.191	0.114	0.060	0.110	0.105	0.117	2.263	2.174	2.397
$\lambda=0.10$	(0.043)	(0.046)	(0.046)	(0.034)	(0.072)	(0.130)	(0.615)	(1.267)	(2.325)
Conservative	0.162	0.107	0.062	0.053	0.054	0.054	1.072	1.077	1.103
$\lambda=0.05$	(0.039)	(0.041)	(0.047)	(0.017)	(0.032)	(0.068)	(0.311)	(0.587)	(1.214)

Comparing Panel B with Panel A, we see that the nonparametric test slightly underestimates the daily timer's ability when the information quality is high. At medium and low information quality, the test delivers quite accurate results. On the other hand, the magnitude of timing is reduced by about two-thirds for the TM and the HM methods, consistent with results by Goetzmann et al. (2000). At medium information quality and medium responsiveness, the average $\hat{\theta}$ for the monthly and daily timer are 0.124 and 0.123, respectively. The average $\hat{\gamma}_{TM}$ changes from 0.101 to 0.031 and the average $\hat{\gamma}_{TM}$ drops from 2.093 to 0.572. Again, the TM and HM measures largely pick up the magnitude of convexity of fund returns vis-à-vis the market return. Linear regressions produce downward (in magnitude) biased estimates when the regressor is measured with error. On the other hand, the nonparametric measure is about the probability of convexity, which is less affected by the measurement frequency.

Finally, we examine the assumption of selectivity being independent of timing. If managers engage in selectivity-based trading that is correlated with the market-wide return, our measure could be biased in unknown ways. To gain some assessment of the impact of selectivity on timing, we apply the timing test on simulated portfolio returns generated by a strategy that selects securities as well as times the market. In each period, the weight on the risky portfolio is determined by the manager's market timing information as before (the rest is invested in 3-month T-bills). On top of that, the manager determines the composition of her risky portfolio according to her information about individual securities. We begin with an equally weighted 100 stock portfolio randomly selected from S&P500 stocks in January 1990. Each month, we replace eight existing stocks with new ones (which represents a 96% annual turnover rate, or about the average turnover rate of those mutual funds in our sample). Such an artificial portfolio is designed to mimic a large-cap mutual fund. We assume the manager has superior selectivity information, so that the dropped stocks are randomly chosen from the bottom quartile of the 100 existing stocks formed by sorting on the next-period stock returns. The added stocks are randomly chosen from the top quartile of the rest of 400 S&P500 stocks. We simulate 1000 series (120 months each) of portfolio returns from such a strategy for each of the nine designs as in Table 1 (Panel A). Results are

Notes to Table 1:

[a] Returns are simulated according to the TM model, with sample sizes equal to 60 and 120 (months). The quality of information is classified into three groups: precise, medium, and coarse, based on the magnitude of σ_η. Similarly, the response intensity is sorted by λ into three groups: aggressive, medium, and conservative. The average estimated market timing parameters using the three methods are displayed and their standard deviations are shown in the parentheses.

[b] We use CRSP S&P500 daily return data from January 1990 to December 1999 as the market return and simulate the return data of a daily timer using the TM model as in Panel A. All return data are then compounded into monthly returns on which we perform the timing test using all three methods. For each design, 1000 simulations are tried and the average estimated parameters are displayed in the table with their standard errors shown in the parentheses.

[c] In this simulation, we add selectivity to market timing. In each period, the weight on the risky portfolio is determined by the manager's market timing information as specified in Table 1, Panel A. Further, the manager determines the composition of her risky portfolio according to her information about individual securities. We begin with an equally weighted 100 stock portfolio randomly selected from S&P500 stocks in January 1990. For the subsequent 119 months, we replace eight existing stocks with new ones each month. The dropped stocks are randomly chosen from the bottom quartile of the 100 existing stocks sorted on the next period returns, and the added stocks are randomly chosen from the top quartile of the rest of 400 S&P500 stocks.

posted in Table 1 (Panel C). When the manager has precise timing information, the nonparametric timing measure is slightly downward biased with the presence of selectivity. Otherwise, results in Table 1 (Panel C) are qualitatively similar to those in Panel A. More importantly, the magnitude of the nonparametric measure varies with information quality, but not much with reaction intensity.

3. Testing the market timing of mutual funds

3.1. Data

In this section, we apply the nonparametric timing test on a large sample of mutual funds. The data in this paper are retrieved from *Morningstar Principia Pro Plus for Mutual Funds* (1980–1999) published by Morningstar in January 2000 and the CRSP Mutual Funds Data. Morningstar offers quality data on surviving funds of all categories on a monthly frequency. To minimize the survivorship bias, we supplement the data set with perished funds from the CRSP Mutual Fund dataset.

In this paper, we focus on diversified domestic equity funds, as do most other studies on mutual fund performance. We also include domestic sector funds that specialize in technology, a sector that was substantially more volatile than the overall stock market during the sample period. It would be interesting to see how well these funds perform in a sector where market timing can be highly rewarding. Morningstar separately records multiple classes of shares issued by the same fund out of basically the same portfolio. We aggregate these classes of shares at the fund level. Therefore, funds in our sample represent unique portfolios. We exclude index funds and enhanced index funds whose managers are not expected to time the market. We also exclude funds that have an R^2 greater than 0.95 from a regression on a best-fitted index, because these funds are highly suspicious of being "closet indexers". Finally, we only use funds (survived or perished) that have at least 2 full years of monthly return data within the 1980–1999 window.[11] Altogether, there are 1827 surviving funds and 110 dead funds in the sample. The sample of dead funds is underrepresented because CRSP does not have data at monthly frequency for about half of the dead funds that perished during the 1980s. Our results are not completely free from the survivorship bias (Carhart, 1997; Carhart et al., 2001).

All returns are expressed in percentage terms. Morningstar reports returns that are computed by taking the change in monthly net asset value (NAV), reinvesting all income and capital gains during the month, and dividing by the starting NAV. Unless otherwise noted, we do not adjust total returns for sales charges (such as front-end or deferred loads and redemption fees). However, the returns do account for management, administrative, 12b-1 fees, and other costs that are automatically taken out of fund assets. We use the monthly return of 3-month T-bills as the proxy for the risk-free rate r_f.[12] To find the proper benchmark for each fund, we regress its returns on four representative indices: S&P500

[11] Requiring funds to stay in business for at least 2 years may subject our results to survivorship bias.

[12] Some studies use a 1-month T-bill rate as the proxy for the risk-free rate. The correlation between the two rates is 0.93 during the sample period.

(benchmark for large-cap funds), S&P400 (for mid-cap funds), Russell 2000 (for small-cap funds), and NASDAQ Composite (for technology funds). The index that gives the largest R^2 is selected as the benchmark for measuring the fund's market timing ability. The average R^2 on a best-fitted index is 82.2% and the average R^2 on the next best-fitted index is 72.4%.

Summary statistics are reported in Table 2. We classify funds into different groups according to their stated prospectus objective and best-fitted indices. The seven objective groups are: small company, growth, equity income, growth and income, asset allocation, balanced, specialty-technology, and aggressive growth. We report the number of funds in each group, the median assets under management, the mean monthly return and the standard deviation, the mean α values from one-factor regressions and their standard deviations, and the mean α values from Fama and French (1993) three-factor regressions and their standard deviations. All together, 26.1% of the live funds and 18.2% of the dead funds beat their benchmark indices before taking out their sales charges.

3.2. Do funds out-guess the market?

In this section, we use the nonparametric measure to test the market timing ability of mutual funds. In Table 3 (Panel A), we report the average market timing $\hat{\theta}$ estimates for all groups of funds, together with the number of funds that have positive and negative $\hat{\theta}$, and the number of funds whose $\hat{\theta}$ values are significantly positive or negative (at 5% and 2.5% significance levels). In computing the average $\hat{\theta}$, we use both equal weighting and standard

Table 2
Summary statistics of mutual funds

Fund group	N	Median assets ($MM)	Mean return	Standard deviation	α_{F1}	Std (α_{F1})	α_{F3}	Std (α_{F3})
By funds prospectus objective								
Small company	333	139.55	1.545	6.042	0.095	0.692	0.079	0.654
Growth	722	187.60	1.744	5.291	− 0.040	0.524	0.007	0.539
Equity income	94	165.40	1.200	3.625	− 0.195	0.290	− 0.198	0.268
Growth and income	256	212.30	1.374	4.154	− 0.189	0.301	− 0.184	0.281
Balanced	188	118.40	1.069	2.760	− 0.118	0.327	− 0.119	0.297
Asset allocation	120	108.80	1.019	2.871	− 0.111	0.234	− 0.108	0.202
Specialty-technology	39	724.40	3.890	9.191	0.711	1.025	1.453	1.214
Aggressive growth	75	119.15	2.063	6.758	− 0.037	0.564	0.189	0.524
By best-fitted indices								
S&P500	803	201.85	1.376	3.861	− 0.177	0.283	− 0.137	0.238
S&P400	240	106.90	1.178	4.431	− 0.156	0.433	− 0.254	0.378
Russell 2000	422	101.70	1.390	5.567	0.257	0.711	− 0.014	0.619
NASDAQ Composite	362	278.90	2.469	6.776	− 0.010	0.619	0.491	0.780
Dead funds	110	–	1.033	4.874	− 0.392	0.532	− 0.420	0.516

There are altogether 1827 live funds and 110 dead ones in the sample. They are actively managed domestic equity funds that have at least 2 years of monthly return data during the sample period from January 1980 to December 1999. N is the number of funds in the group. Median assets are expressed in terms of million dollars. Mean return and standard deviation are the sample mean and standard deviation of monthly returns. α_{F1} is the mean α value from one-factor regressions on the best-fitted indices and Std(α_{F1}) is its sample standard deviation. α_{F3} is the mean α value from the Fama and French (1993) three-factor regressions and Std(α_{F3}) is its sample standard deviation.

Table 3
Market timing of mutual funds

Panel A: the nonparametric method[a]

Fund group	Mean $\hat{\theta}$ (%)	Weighted average $\hat{\theta}$ (%)	# $\hat{\theta}$>0 ($\hat{\theta}$< 0)	% $\hat{\theta}$>0 ($\hat{\theta}$< 0) significant at 5%	% $\hat{\theta}$>0 ($\hat{\theta}$< 0) significant at 2.5%
By funds prospectus objective					
Small company	– 1.766 (0.356)	– 1.389 (0.289)	133 (200)	2.40 (9.31)	0.60 (4.81)
Growth	– 0.725 (0.245)	– 0.654 (0.184)	320 (402)	4.57 (5.96)	1.80 (3.46)
Equity income	– 2.658 (0.592)	– 3.226 (0.448)	21 (73)	2.13 (18.1)	1.06 (13.8)
Growth and income	– 1.549 (0.357)	– 1.764 (0.266)	90 (166)	3.52 (10.94)	1.95 (5.47)
Balanced	– 1.924 (0.442)	– 2.245 (0.331)	61 (127)	1.06 (12.23)	1.06 (8.51)
Asset allocation	– 1.896 (0.634)	– 1.775 (0.501)	45 (75)	0.83 (5.00)	0.00 (3.33)
Specialty-technology	2.633 (1.110)	2.494 (0.856)	28 (11)	7.69 (2.56)	5.13 (0.00)
Aggressive growth	– 2.393 (0.695)	– 2.368 (0.541)	21 (54)	1.33 (10.67)	0.00 (5.33)
Asset allocation, balanced, and growth and income	– 1.748 (0.257)	– 1.927 (0.192)	196 (367)	2.13 (10.10)	1.24 (6.03)
By best-fitted indices					
S&P500	– 0.163 (0.219)	– 0.733 (0.159)	360 (443)	4.86 (8.22)	2.12 (5.11)
S&P400	– 4.368 (0.417)	– 3.725 (0.313)	50 (190)	0.83 (16.3)	0.83 (10.83)
Russell 2000	– 0.699 (0.319)	– 0.294 (0.259)	203 (219)	2.61 (3.79)	0.47 (1.90)
NASDAQ Composite	– 2.624 (0.341)	– 2.444 (0.270)	106 (256)	1.93 (9.94)	1.10 (4.70)
Dead funds	– 2.621 (0.672)	– 2.932 (0.583)	41 (69)	3.63 (19.03)	1.82 (12.73)

Panel B: conditional market timing[b]

Fund group	Mean $\tilde{\theta}$ (%)	Weighted average $\tilde{\theta}$ (%)	# $\tilde{\theta}$>0 ($\tilde{\theta}$< 0)	% $\tilde{\theta}$>0 ($\tilde{\theta}$< 0) significant at 5%	% $\tilde{\theta}$>0 ($\tilde{\theta}$< 0) significant at 2.5%
By funds prospectus objective					
Small company	0.688 (0.340)	– 0.161 (0.267)	165 (168)	2.40 (5.11)	1.50 (2.40)
Growth	1.025 (0.253)	0.143 (0.181)	386 (336)	1.80 (2.91)	0.83 (1.39)
Equity income	– 2.670 (0.591)	– 2.528 (0.453)	20 (74)	1.06 (15.96)	0.00 (8.51)
Growth and income	– 1.166 (0.359)	– 1.200 (0.260)	97 (159)	0.78 (8.59)	0.00 (5.08)
Balanced	– 1.094 (0.435)	– 1.395 (0.317)	62 (126)	0.53 (12.23)	0.00 (6.39)
Asset allocation	– 0.755 (0.600)	– 1.248 (0.469)	44 (76)	0.83 (7.50)	0.83 (5.83)
Specialty-technology	3.086 (1.104)	1.770 (0.839)	26 (13)	12.82 (5.13)	12.82 (2.56)
Aggressive growth	1.949 (0.677)	0.833 (0.506)	47 (28)	5.33 (2.67)	5.33 (2.67)
Asset allocation, balanced, and growth and income	– 1.054 (0.253)	– 1.274 (0.185)	203 (361)	0.71 (9.57)	0.18 (5.67)
By best-fitted indices					
S&P500	– 0.503 (0.229)	– 0.852 (0.159)	348 (455)	0.37 (7.22)	0.00 (4.36)
S&P400	– 0.946 (0.375)	– 1.209 (0.289)	84 (156)	0.42 (9.58)	0.42 (4.58)
Russell 2000	0.474 (0.308)	– 0.471 (0.241)	195 (227)	1.90 (5.21)	1.42 (2.84)
NASDAQ Composite	2.267 (0.342)	1.046 (0.258)	220 (142)	6.35 (2.21)	3.87 (0.83)
Dead funds	– 3.862 (0.761)	– 3.111 (0.611)	31 (79)	0.00 (20.91)	0.00 (13.64)

error weighting. The latter assigns a weight to each fund's $\hat{\theta}$ that is inversely proportional to its standard error.

Overall, there is no evidence that mutual fund managers possess superior market timing abilities. The average $\hat{\theta}$ value of live funds is -1.33% and that of the dead funds is -2.62%. As expected, managers of dead funds are less successful market timers. According to the interpretation of θ as defined in Eq. (2), the probability that the manager of an average live fund moves the fund's exposure to the market in the correct direction is 1.33 percentage points lower than the probability of a move in the wrong direction. The total numbers of funds with positive and negative $\hat{\theta}$ values are 719 and 1108. Out of eight fund groups by stated prospectus objectives, only the special-technology fund group shows up an average $\hat{\theta}$ parameter above the neutral level zero. All other groups have negative average timing coefficients. They also have more individual funds with $\hat{\theta}$ values significantly (at both 5% and 2.5% significance levels) different from zero on the negative side than on the positive side. All fund groups by best-fitted indices have average $\hat{\theta}$ values below zero. This slightly perverse market timing ability is also found in Ferson and Schadt (1996), Becker et al. (1999), Edelen (1999), and Goetzmann et al. (2000).

The top 5% of market timers have $\hat{\theta}$ values above 8.47%. Farnsworth et al. (in press), through a simulation procedure, find that the best performing mutual funds (those in the upper 5%) have performance similar to artificial mutual funds that have correlation values between manager's signal and the subsequent market return ranging from 0.24 to 0.32. To make the two results comparable, we derive the one-to-one relationship between θ and the correlation coefficient in the TM model. Assume market returns are *i.i.d.* normal with variance σ_m^2. Numerically, the relationship between the value of θ and the correlation $\mathrm{corr}(y_t, r_{m,t+1}) = \frac{\sigma_m}{\sqrt{\sigma_m^2+\sigma_\eta^2}} = \frac{1}{\sqrt{1+\sigma_\eta^2/\sigma_m^2}}$ is as follows:

$\mathrm{corr}(y_t, r_{m,t+1})$	0.0	0.1	0.2	0.3	0.5	1.0
θ (%)	0.0	6.4	12.8	19.4	33.3	100.0

If a manager makes correct predictions (in terms of the relative ranking of market returns) 20% more often than making wrong ones, the correlation between her private signal and the subsequent market return is about 0.3. θ equal to 8.47% corresponds to a correlation

Notes to Table 3:

[a] This panel reports the nonparametric timing measures of mutual funds. The second column lists the equally weighted average $\hat{\theta}$ of all funds within the group in percentage points. Its standard error is reported in the parenthesis. The weighted average $\hat{\theta}$ (and its standard error in the parenthesis) is reported in the third column, using weights that are inversely proportional to each estimate's standard error. The fourth column shows the number of funds that have $\hat{\theta}$ above (below) zero. The fifth and sixth columns report the percentage of funds in each group that have $\hat{\theta}$ above (below) zero at 5% and 2.5% significance level.

[b] In this panel, $\tilde{\theta}$ is calculated using Eq. (12). The residual market returns are the residuals of each index returns from a regression on lagged instrumental variables, which include (i) the de-trended 1-month T-bill yield; (ii) the dividend-to-price ratio for the CRSP value-weighted NYSE and Amex stock indices; (iii) the slope of the U.S. Treasury yield curve measured as the difference between the 4- and 1-year fixed maturity bond yields; (iv) a dummy variable for the month of January; and (v) the index's own lagged return. The residual fund returns are the residuals of each fund returns from a regression on the same lagged instrumental variables (i)–(iv), plus the fund's own lagged return.

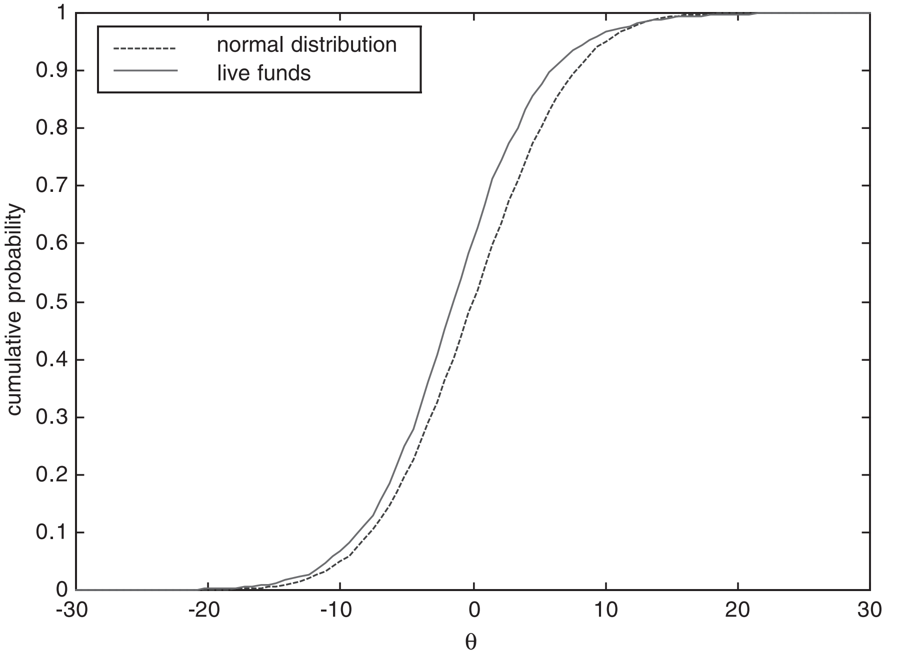

Fig. 1. Distribution of market timing of live funds. To check whether superior ability exists among fund managers beyond the tail probability of a distribution, this graph plots the empirical cumulative distribution of live funds' $\hat{\theta}$ values against a reference normal distribution that has the same standard deviation but centers on zero (no ability). The empirical distribution lies all the way above the normal distribution, suggesting that the distribution of live mutual funds' market timing ability is first-order stochastically dominated (FOSD) by a comparable distribution. The FOSD hypothesis passed the Klecan et al.'s (1991) test at a 1% significance level.

between the manager's signal and the future market return being around 0.14. Our estimate of the timing skills of mutual fund managers is lower than that of Farnsworth et al. (in press).

Since $\hat{\theta}$ is a random variable, we are bound to find funds with significantly positive or negative $\hat{\theta}$ values even when the truth is $\theta = 0$ because of the large sample size. Hence, we would like to see whether superior ability exists among fund managers beyond what is implied by the tail probability of a normal distribution. Figs. 1 and 2 plot the empirical cumulative distributions of $\hat{\theta}$ of all live and dead funds against normal distributions with the same standard deviation but centered on the neutral value of zero (no ability). The empirical distributions lie mostly above the reference normal distributions, suggesting that the empirical distributions are first-order stochastically dominated by their respective reference normal distributions based on the hypothesis of no ability. We applied the Klecan et al.'s (1991) robust test for stochastic dominance[13] on both empirical distributions of $\hat{\theta}$

[13] The test is based on the following idea: if F does not first-order stochastically dominate G, then max $d = \max[F(x) - G(x)] > 0$, for all values of x within the support. The test statistic is built on the empirical analogue $\hat{d}$ by constructing max $\hat{d}(x)$ of F against G at fine grids. The resulting statistic follows a non-standard distribution and its standard error is obtained through an algorithm provided in Klecan et al. (1991).

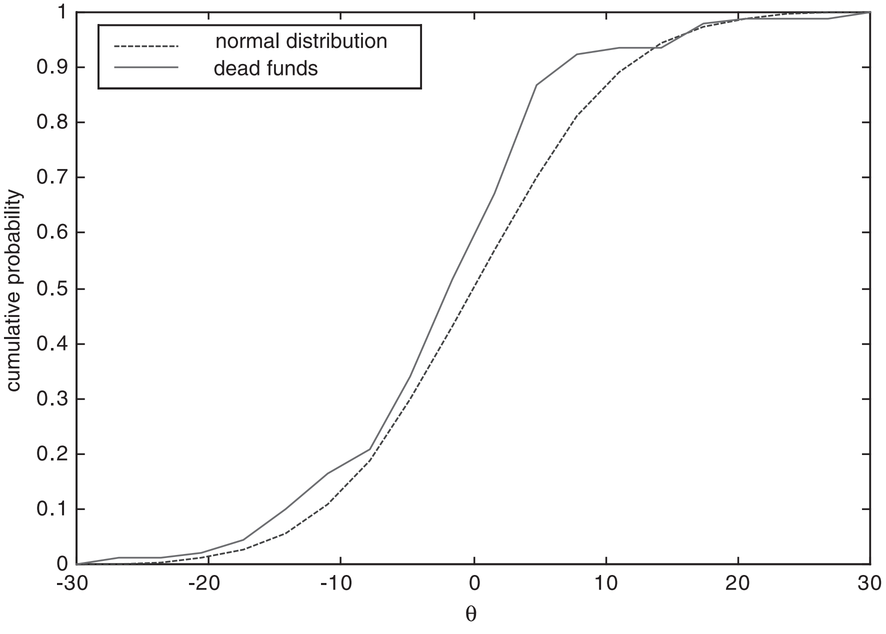

Fig. 2. Distribution of market timing of dead funds. This graph plots the empirical cumulative distribution of dead funds' $\hat{\theta}$ values against a reference normal distribution that has the same standard deviation but centers on zero (no ability). The empirical distribution lies mostly above the normal distribution, suggesting that the distribution of dead mutual funds' market timing ability is FOSD by a comparable distribution. The FOSD hypothesis passed the Klecan et al.'s (1991) test at 5% significance level.

and their respective normal distributions with equal variances. The test rejects the hypothesis that the normal distribution does not first-order stochastically dominate the empirical distribution of live funds at the 1% significance level. Dead funds are rejected at the 5% significance level (the significance level is higher due to a smaller sample size). This result implies that the distribution of the timing ability of actively managed funds is first-order stochastically dominated by a normal distribution that would prevail if no superior ability exists on average and performance is similarly dispersed. If an investor does not have superior knowledge about individual managers' ability, she would be better off in terms of market timing choosing an index fund (no timing) than randomly choosing a fund from the pool of actively managed funds.

As a comparison, we also compute the TM and HM measures of market timing for the same sample (not tabulated). The correlation coefficients between the TM and the HM estimates are above 0.90 for all groups. The correlation coefficients between the nonparametric and the TM estimates vary from 0.29 to 0.62, while those between the nonparametric and the HM measures range from 0.44 to 0.73. The latter correlation is higher because, as we discussed in Section 2, the HM measure is more sensitive to both information quality and response intensity, while the TM measure basically reflects the latter. The group of special-technology funds has positive average market timing using all three methods. All other fund groups except the asset allocation group have negative average timing parameters using all three methods. The average $\hat{\gamma}_{TM}$ value of asset allocation funds is positive, but it is very close to zero (in a magnitude of 0.13 of its standard error). The aggregate picture of mutual funds' market timing performance is similar across the different methods. At individual funds level, however, the three methods contain different information.

The market timing performance of mutual funds as measured here is subject to several biases. The first is the survivorship bias. The number of mutual funds in the US grew dramatically in the 1990s. About 42% of the surviving funds in our sample are less than 5 years old as of 1999. While we try our best to include perished funds during the period, they are still underrepresented in the sample and the results are not completely free from the survivorship bias (Carhart et al., 2001). However, if managers of perished funds do not have better market timing ability than those of live funds, then our evidence of no significant market timing ability is likely to hold up in a sample with full information about perished funds.

The second potential bias comes from securities selectivity. Not all equity funds are explicit market timers. When a manager's information about individual securities (or industries) is correlated with her information about market trends, selectivity may bias market timing inferences in unknown ways. To address this issue, we look at a sub-group of funds that are more likely to be market timers, namely funds in the asset allocation, balanced, and growth-and-income groups. That adds up to 564 funds. The market timing performance of these three groups of funds (see Table 3 (Panel A)) is very similar to the aggregate performance of all actively managed domestic equity funds. Further, the group of funds that is most expected to time the market—asset allocation funds—has an average $\hat{\theta}_n$ of -1.8%, not indicative of superior market timing ability.

Third, the results can be biased if fund managers time the market based on conditional information. The conditional performance evaluation literature says that investors will neither pay managers to use readily available public information nor penalize managers for mistiming that is predictable from public information. According to this view, any effect of timing or mistiming stemming from the public information component is spurious timing. Work by Ferson and Schadt (1996), Ferson and Warther (1996), and Becker et al. (1999) show that fund managers do vary β according to publicly available information, such as past returns, dividend yields, and the term structure. More importantly, their results indicate that conditioning β on such public information removes part of the negative value of the $\hat{\gamma}_{TM}$ coefficient in a TM-type regression.

To incorporate conditional information into our test, we use Eq. (12) to analyze mutual funds' ability to time the part of market returns that is not predicted by public information. To represent public information, we use a collection of variables that are adopted by previous studies on conditional market timing (Becker et al., 1999; Ferson and Khang, 2001). The variables are as follows: (1) the lagged level of the 1-month Treasury bill yield, less its 12-month lagged moving average; (2) the lagged dividend-to-price ratio for the CRSP value-weighted NYSE and Amex stock indices; (3) the lagged slope of the US Treasury yield curve, measured as the difference between the 4- and 1-year fixed maturity bond yields from the CRSP Fama–Bliss files; and (4) a dummy variable for the month of January. We also add an own lagged return term of the index (or of the fund). We use 2 months of lagged values for the first three variables. The average $\tilde{\theta}$ measure (as defined in

Eq. (12)) of all funds is 0.21% (standard error 0.15%); the standard error weighted average of $\tilde{\theta}$ is -0.46% (standard error 0.11%). Altogether, 847 (980) funds have $\tilde{\theta}$ estimates greater (smaller) than zero. Results of conditional timing performance sorted by fund groups are reported in Table 3 (Panel B). Similar to findings in Ferson and Schadt (1996) and Becker et al. (1999), perverse market timing largely vanishes once we filter out public information from the returns of both funds and indices. Furthermore, after controlling for the public information, we find no evidence that mutual funds have market timing ability based on superior information.

Finally, we want to check that the deviation of the timing measure from the neutral level of zero is mainly due to timing instead of being due to holding particular classes of assets that exhibit timing characteristics. Jagannathan and Korajczyk (1986) predict that funds investing in small stocks can show spurious timing against a market benchmark that consists mainly of big stocks. For example, they find that an equally weighted stock index shows "timing" relative to a value-weighted index. This is because small stocks have payoffs resembling that of a call option related to its underlying assets. This is why we classify mutual funds into groups by best-fitted indices and test market timing of funds relative to their own benchmarks, instead of using a uniform market portfolio. To check whether common factors (such as the book-to-market ratio) may interfere with our timing test results, we apply the same nonparametric test to the Fama and French (1993) factor portfolios against the market indices for the same sample period. Although the portfolios actually held by mutual funds may not resemble the hypothetical Fama–French portfolios, such a test gives us some guidance on how investment style based on size and book-to-market can affect timing performance. Altogether, there are six factor-mimicking portfolios sorted by size (small and big) and book-to-market equity (low, medium, and high). We find none of the six portfolios exhibits timing characteristics relative to the indices we use at less than 10% significance level. Therefore, our results on market timing should not be driven by the characteristics of the portfolios that mutual funds hold. However, the TM and HM regressions show that the small stock portfolios sorted by book-to-market exhibit some spurious timing effect relative to the small-cap index Russell 2000. Both regressions produce significantly positive (at a 5% level) timing coefficients for the small/low portfolio, and significantly negative (at a 5% level) timing coefficients for the small/medium and small/high portfolios. This illustrates that performance measurement can be sensitive to benchmark specification.

3.3. Some related questions

In this section, we analyze some related questions about market timing by focusing on the live funds. The issues of interest are: (1) Do experienced managers do better in timing the market? (2) Is it easier for small funds to time the market? (3) Can the higher turnover funds better time the market? (4) Is market timing impaired by the in-and out-flows of investment money? In Table 4, we compare the timing performance of mutual funds sorted by fund characteristics related to the aforementioned issues. Overall, the relationship between the average timing performance and fund characteristics is weak, so it is difficult to predict market timing ability of fund managers from observable characteristics.

Table 4
Market timing: related issues

Fund characteristics	# Funds	Weighted average $\hat{\theta}$ (%)	Standard error (%)	% $\hat{\theta}>0$ significant at 5%	% $\hat{\theta}<0$ significant at 5%	% $\hat{\theta}>0$ significant at 2.5%	% $\hat{\theta}<0$ significant at 2.5%
Panel A: fund age (years since inception)							
< 5	768	− 1.533	0.282	2.53	11.09	0.97	7.19
≥ 5 and < 10	545	− 1.315	0.206	4.77	11.38	2.20	6.97
≥ 10	514	− 1.250	0.137	2.60	4.94	1.04	2.21
Panel B: manager tenure (years)							
< 3	386	− 1.550	0.251	3.89	9.84	1.81	6.73
≥ 3 and < 5	741	− 1.284	0.221	2.43	6.07	0.67	2.97
≥ 5	690	− 1.326	0.150	3.77	10.72	1.88	6.38
Panel C: fund size ($mm)							
< 20	362	− 0.629	0.331	4.70	5.24	2.49	3.31
≥ 20 and < 100	528	− 1.420	0.239	2.27	7.95	0.95	4.92
≥ 100 and < 500	515	− 1.432	0.207	3.50	9.51	1.55	5.63
≥ 500	420	− 1.635	0.179	2.86	11.19	0.71	5.95
Panel D: turnover ratio (annual %)							
All funds							
< 50	522	− 1.419	0.197	3.43	8.32	2.29	5.06
≥ 50 and < 100	613	− 1.130	0.184	4.02	8.55	1.47	4.73
>100	573	− 1.580	0.210	2.44	10.15	0.52	6.13
Asset allocation funds only							
< 50	42	− 2.016	0.805	2.38	4.76	0.00	4.76
≥ 50 and < 100	26	− 0.878	0.974	0.00	0.00	0.00	0.00
>100	44	− 2.565	0.868	0.00	9.09	0.00	4.55
Panel E: load charges							
No-load funds	1012	− 1.467	0.154	3.06	8.50	0.99	5.43
Load funds	815	− 1.215	0.166	3.44	8.71	1.84	4.54
Panel F: minimum initial purchase (US$1000)							
< 25	1557	− 1.330	0.121	3.21	8.41	1.54	4.95
≥ 25	210	− 1.822	0.392	2.86	10.95	0.48	7.14

In this table, we compare the timing performance of mutual funds sorted by fund characteristics. In each row, we report the number of funds, the inverse standard error weighted average $\hat{\theta}$ and its standard error, and the percentage of funds that outperform or underperform the market at 5% and 2.5% significance levels. Fund age and manager tenure are proxies for manager experience. The median age and manager tenure are 5 and 4 years, respectively. The median fund size is US$104.75 million. The turnover ratio is the average annual rate during the sample period. The median turnover ratio is 69.0%. Load charges and minimum initial purchase are proxies for fund flow stability. Load charges are the sum of front and deferred loads. The 55.3% of the funds in the sample are no-load funds. The 54.1% of the funds have minimum initial purchase of US$1000 or less. The 11.5% of the funds have minimum initial purchase of US$25,000 or more.

3.3.1. Does experience matter?

If established funds are more likely to have experienced managers, we can test whether fund age contributes to better timing ability. Results are posted in Table 4 (Panel A). Estimated average market timing ability increases monotonically with the age of funds.

Young funds (less than 5 years old) have a weighted average $\hat{\theta}$ of -1.533%. The same estimates for the medium-aged funds (between 5 and 10 years old) and old funds (more than 10 years old) are -1.315% and -1.250%, respectively. Therefore, on average, older funds are doing better than younger funds. However, the survivorship bias in the sample would also work in favor of the older funds.

We can also use manager tenure directly as a proxy for experience. In the data, we only have the tenure information of the current managers. Accordingly, we crop out the return data that they are responsible for. We divide all managers into three groups, depending on whether their tenure is less than 3 years, between 3 and 5 years, or 5 years and more. The results are shown in Table 4 (Panel B). The least experienced manager group (manager tenure less than 3 years) has an average $\hat{\theta}$ of -1.550%, those of more experienced groups (manager tenure between 3 and 5 years, and 5 years or more) are -1.284% and -1.326%, respectively. And there are proportionately more out of experienced managers who turn in extraordinary records. However, the length of manager tenure is endogenous to mutual fund performance. The manager who remains in the position for a long time is likely to have produced a reasonably good record. Furthermore, the differences are not statistically significant.

3.3.2. Do small funds fare better?

One may argue that small funds are in a better position to time the market since it is easier for them to buy and sell quickly without affecting market prices. According to the hypothesis of efficient market with costly information, activities out of superior information must be "small" relative to the market to earn superior returns (Mayers and Rice, 1979; Grossman and Stiglitz, 1980).

The size of the funds varies from less than US\$1 million to US\$99,184 million in the sample. The median is US\$105 million. We divide the funds into four groups: micro (under US\$20 million), small (up to US\$100 million), big (up to US\$500 million), and huge (US\$500 million or more). Results are shown in Table 4 (Panel C). Estimated market timing ability deteriorates monotonically with fund size. Estimated average $\hat{\theta}$ values are -0.629%, -1.420%, -1.432%, and -1.635% from the smallest to the biggest fund groups. The micro fund group beats the huge fund group at 5% significance. It seems that, on average, small funds are doing better than their larger counterparts. There is a larger proportion of mis-timers from big and huge funds, who drive down the group average results.

Chevalier and Ellison (1999) document that small mutual funds are often managed by more experienced managers. Too see whether part of the small fund timing premium is attributed to manager experience, we make a comparison by a two-way sort on fund size and manager tenure. It turns out that the difference in timing performance due to fund size is similar across manager tenure. For managers with less than 3 years in tenure, the timing performance of micro/small funds and big/huge funds are -1.335% versus -1.840%. For managers with 5 years or more of tenure, the corresponding estimates are -1.044% and -1.478%. The differences are not statistically significant.

3.3.3. Is high turnover rate justified as timing?

The turnover rate of a fund is a proxy for how frequently a manager trades her portfolio. The inverse of a fund's turnover rate is the average holding period for a security in that fund. If one maintains an S&P500 benchmark portfolio, like the Vanguard 500 Index Fund, the average annual turnover rate is about 4–6% for the past 10 years. The average turnover rate of actively managed funds investing in the same market is 92.8%. If the turnover rate is positively correlated with the frequency of timing-oriented trading, we would like to see to what extent high turnover represents successful market timing.

Out of 1827, 1708 report their annual turnover rates to the Morningstar database. We divide all reporting funds into three categories according to their average annual turnover rates during the sample period: low (less than 50%), medium (between 50% and 100%), and high (100% or higher) turnover. Results are shown in Table 4 (Panel D). It turns out that the highest turnover fund group has the worst market timing record, with an average $\hat{\theta}$ value of − 1.580%. Moderate turnover funds (− 1.130%) slightly outperform low turnover funds (− 1.419%). This result is consistent with Morningstar's report that mutual funds with annual turnover higher than 100% significantly underperform their lower turnover counterparts.[14] In particular, the group of asset allocation funds usually explicitly market themselves as market timers. These managers often use a flexible combination of stocks, bonds, and cash. They also shift assets frequently based on their analyses of market trends. The average annual turnover rate of asset allocators is 114.4%, which is higher than the average of all funds. The relationship between average timing and turnover rate for asset allocation funds is very similar to that of other funds. Results are shown in Table 4 (Panel B). Asset allocators with moderate turnovers have the best timing performance (− 0.878%), outperforming that of low turnover asset allocators (− 2.016%) and high turnover ones (− 2.565%). However, the differences are not statistically significant.

High turnover rate may result from frequent trading that is not related to market timing. Dow and Gorton (1997) consider a model where portfolio managers trade, even though they have no reason to, because their clients cannot distinguish "actively doing nothing" from "simply doing nothing". Lakonishok et al. (1991) tell a story that fund managers dress up their portfolios (by selling off the losers) before disclosing to the public in order to make the composition look "smart". Haugen and Lakonishok (1988) suggest window dressing by professional money managers as a possible explanation of the "January Effect". Funds engaging in such window dressing activities are selling to avoid apologizing for and defending a loser stock's presence to clients even though it is bad timing to sell (Lakonishok et al., 1991). In equilibrium, such trading activities, on average, lose to more informed traders and would appear as buying/selling at inopportune times, or mis-timing.

3.3.4. Do investor flows affect market timing?

One plausible explanation for mutual funds' unsatisfactory timing performance is the funds' open-ended nature (as opposed to closed-end funds). While fund managers try to time the market, there are investors who attempt at timing the mutual funds. When the market fares well, new money flows in and lowers the portfolio β by increasing the portfolio's cash position. Conversely, outflows increase β when the market experiences a downturn. Additionally, large redemption orders can force funds to liquidate shares in a falling market. From this point of view, market mis-timing of mutual funds constitutes a price that investors have to pay for the liquidity that they enjoy with open-end funds.

[14] Source: Morningstar Report, September 12, 1997.

Ferson and Warther (1996) and Edelen (1999) document a negative relation between a fund's risk-adjusted return and investor flows. They attribute the negative return performance to the cost of liquidity-motivated trading.

Such logic implicitly assumes that fund investors can time the market ahead of the fund managers. Only if investors' money flows in prior to market ascendancy or flows out prior to market descent will investor flows offset fund managers' market timing endeavor. Gruber (1996) and Zheng (1999) provide some evidence that funds receiving more money subsequently beat the market—the "smart money" effect—but in the aggregate such effect is weak. Warther (1995) documents a positive relation between flows and subsequent returns in weekly data. Short-term switchers in and out of funds are more likely to attack on no-load funds, thus taking advantage of the cost-free entry and exit. Hence, we look at possible differences in timing between load and no-load funds. Out of the 1827 funds, 1012 are no-load funds. Results in Table 4 (Panel E) show that load funds (with an average $\hat{\theta}$ value of -1.215%) slightly outperform no-load funds (-1.467%), but overall the timing performance profiles of the two groups are very similar. Further, from the investors' point of view, index funds also provide the liquidity service, but the market timing cost of such service has been negligible.[15]

There is evidence in the literature that retail investors and institutional investors differ, both with respect to the timing and to the information content of their investments. By comparing the return autocorrelation of securities and portfolios dominated by institutional investors with those dominated by individual investors, Sias and Starks (1997) find that institutional investors are more likely to be informed traders. If informative signals about market returns contain a market-wide component (i.e., they are cross-sectionally correlated), then funds mainly open to institutional investors are likely to have more difficulty timing the market because the informed in-and out-flows of investor money can offset the funds' attempts to time the market. To see whether this effect exits in data, we divide funds into retail and institutional funds using a US$25,000 minimum initial investment as the cutoff between the two. In Table 4 (Panel F), we compare the timing skills of the two groups of funds. There are 1557 retail funds and 210 institutional funds in the sample. Retail funds have an average $\hat{\theta}$ of -1.330%, while that of the institutional funds is -1.822%. The result is consistent with the hypothesis that the funds' market timing is impaired by informed investor money flows, but the difference is not significant.

4. Conclusion

In this paper, we propose a nonparametric test for market timing ability and apply the analysis to a large sample of domestic equity funds that have different benchmark indices. Overall, we do not find superior timing abilities among mutual funds. In particular, the number of funds that display extraordinary timing ability is smaller than the right-tail probability of the reference normal distribution that centers on zero with similar dispersion.

[15] Since its inception, for example, the Vanguard Total Stock Market Index Fund has lagged the index return by only 0.3%, the Vanguard 500 Index Fund about 0.2% annually, although they stand ready for investors' purchase or redemption as open-end funds.

We find that the average timing performance bears a positive relation with fund age or management tenure (proxy for manager experience) and fund load (proxy for the stability of fund flow). It is negatively related to fund size. Further, funds with moderate turnover rates outperform both low and high turnover funds. However, overall, the relation between market timing ability and fund characteristics is very weak. The differences of average market timing abilities between different fund groups are too small to have much economic significance. This implies that it is difficult for investors to choose good timers from the universe of mutual funds based on commonly observable characteristics.

Acknowledgements

I thank Jason Abrevaya, Gene Amromin, Brad Barber, Judy Chavelier, Douglas Diamond, Jie Gan, Geert Bekaert (the editor), Lars Hansen, Milton Harris, Terry Odean, Zhenyu Wang, an anonymous referee, and workshop participants at the University of Chicago, the University of California (Davis), and AFA 2002 meeting for their helpful comments. I am especially indebted to Wayne Ferson for extensive and constructive comments and suggestions on earlier drafts. All errors are mine.

References

Abrevaya, J., Jiang, W., 2001. Pairwise-slopes statistics for testing curvature. Working paper, University of Chicago Graduate School of Business and Columbia Business School.

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

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

Barber, B., Odean, T., 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. Journal of Finance 55, 773–806.

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

Bollen, N., Busse, J., 2001. On the timing ability of mutual fund managers. Journal of Finance 56, 1075–1094.

Breen, W., Jagannathan, R., Ofer, A., 1986. Correcting for heteroscedasticity in tests for market timing ability. Journal of Business 59, 585–598.

Breen, W., Glosten, L., Jagannathan, R., 1989. Economic significance of predictable variations in stock index returns. Journal of Finance 44, 1177–1189.

Busse, J., 1999. Volatility timing in mutual funds: evidence from daily returns. Review of Financial Studies 12, 1009–1041.

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

Carhart, M., Carpenter, J., Lynch, A., Musto, D., 2001. Mutual fund survivorship. Review of Financial Studies 15, 1439–1463.

Carpenter, J., 2000. Does option compensation increase managerial risk appetite? Journal of Finance 55, 2311–2331.

Chevalier, J., Ellison, G., 1997. Risk taking by mutual funds as a response to incentives. Journal of Political Economy 105, 1167–1200.

Chevalier, J., Ellison, G., 1999. Are some mutual funds managers better than others? Cross-sectional patterns in behavior and performance. Journal of Finance 54, 875–899.

Coggin, D., Fabozzi, F., Rahman, S., 1993. The investment performance of U.S. equity pension fund managers: an empirical investigation. Journal of Finance 48, 1039–1055.

Dow, J., Gorton, G., 1997. Noise trading, delegated portfolio management, and economic welfare. Journal of Political Economy 105, 1024–1050.
Edelen, R., 1999. Investor flows and the assessed performance of open-end mutual funds. Journal of Financial Economics 53, 439–466.
Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.
Farnsworth, H., Ferson, W., Jackson, D., Todd, S., 2001. Performance evaluation with stochastic discount factors. Journal of Business (in press).
Ferson, W., Khang, K., 2001. Conditional performance measurement using portfolio weights: evidence for pension funds. Journal of Financial Economics 65, 249–282.
Ferson, W., Schadt, R., 1996. Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51, 425–461.
Ferson, W., Warther, V., 1996. Evaluating fund performance in a dynamic market. Financial Analysts Journal 52, 20–28.
Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.
Goetzmann, W., Ingersoll Jr., J., Ivkovich, Z., 2000. Monthly measurement of daily timers. Journal of Financial and Quantitative Analysis 35, 257–290.
Graham, J., Harvey, C., 1996. Market timing ability and volatility implied in investment newsletters' asset allocation recommendations. Journal of Financial Economics 42, 397–421.
Grinblatt, M., Titman, S., 1989. Portfolio performance evaluation: old issues and new insights. Review of Financial Studies 2, 393–421.
Grossman, S., Stiglitz, J., 1980. On the impossibility of informationally efficient markets. American Economics Review 70, 393–408.
Gruber, M., 1996. Another puzzle: the growth in actively managed mutual funds. Journal of Finance 51, 783–809.
Haugen, R., Lakonishok, J., 1988. The Incredible January Effect. Dow Jones Irwin, Homewood.
Henriksson, R., Merton, R., 1981. On market timing and investment performance: II. Statistical procedures for evaluating forecasting skills. Journal of Business 54, 513–533.
Jagannathan, R., Korajczyk, R., 1986. Assessing the market timing performance of managed portfolios. Journal of Business 59, 217–235.
Jensen, M., 1972. Optimal utilization of market forecasts and the evaluation of investment portfolio performance. In: Szego, G., Shell, K. (Eds.), Mathematical Methods in Investment and Finance. North Holland, Amsterdam, pp. 310–335.
Klecan, L., McFadden, R., McFadden, D., 1991. A robust test for stochastic dominance. Unpublished paper, MIT.
Kosik, J., Pontiff, J., 1999. How are derivatives used? Evidence from the mutual fund industry. Journal of Finance 54, 791–816.
Kothari, S., Warner, J., 2001. Evaluating mutual fund performance. Journal of Finance 56, 1985–2010.
Lakonishok, J., Shleifer, A., Thaler, R., Vishny, R., 1991. Window dressing by pension fund managers. American Economic Review Papers and Proceedings 81, 227–231.
Laplante, M., 2001. Conditional market timing with heteroskedasticity. Working paper, University of Washington.
Mayers, D., Rice, E., 1979. Measuring portfolio performance and the empirical content of asset pricing models. Journal of Financial Economics 7, 3–29.
Odean, T., 1998. Are investors reluctant to realize their losses? Journal of Finance 53, 1775–1798.
Serfling, R., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.
Sias, R., Starks, L., 1997. Return autocorrelation and institutional investors. Journal of Financial Economics 46, 103–131.
Treynor, J., Mazuy, K., 1966. Can mutual funds outguess the market? Harvard Business Review 44, 131–136.
Warther, V., 1995. Aggregate mutual fund flows and security returns. Journal of Financial Economics 39, 209–235.
Zheng, L., 1999. Is money smart? A study of mutual fund investors' fund selection ability. Journal of Finance 54, 901–933.