THE JOURNAL OF FINANCE • VOL. LVI, NO. 3 • JUNE 2001

On the Timing Ability of Mutual Fund Managers

NICOLAS P. B. BOLLEN and JEFFREY A. BUSSE*

ABSTRACT

Existing studies of mutual fund market timing analyze monthly returns and find little evidence of timing ability. We show that daily tests are more powerful and that mutual funds exhibit significant timing ability more often in daily tests than in monthly tests. We construct a set of synthetic fund returns in order to control for spurious results. The daily timing coefficients of the majority of funds are significantly different from their synthetic counterparts. These results suggest that mutual funds may possess more timing ability than previously documented.

THE PERFORMANCE OF MUTUAL FUNDS RECEIVES a great deal of attention from both practitioners and academics. Almost 50 percent of U.S. households invest in mutual funds, with an aggregate investment of over five trillion dollars (Investment Company Institute, 2000). Given the size of their stake, the investing public's interest in identifying successful fund managers is understandable, especially in light of mounting evidence that the returns of most actively managed funds are lower than index fund returns.[1] From an academic perspective, the goal of identifying superior fund managers is interesting because it challenges the efficient market hypothesis.

In this paper, we examine the ability of mutual fund managers to time the market, that is, to increase a fund's exposure to the market index prior to market advances and to decrease exposure prior to market declines. Most existing studies find little evidence that fund managers possess market timing ability. Treynor and Mazuy ((1966), hereafter referred to as TM), for example, develop a test of market timing and find significant ability in only 1 fund out of 57 in their sample. Henriksson (1984) uses the market timing test of Henriksson and Merton ((1981), hereafter referred to as HM) and finds that only 3 funds out of 116 exhibit significant positive market timing ability. Graham and Harvey (1996) analyze investment newsletters' suggested allocations between equity and cash, thereby measuring explicitly the

* Bollen is Assistant Professor of Finance at the David Eccles School of Business, University of Utah. Busse is Assistant Professor of Finance at the Goizueta Business School, Emory University. The authors thank René Stultz, an anonymous referee, Uri Loewenstein, Tom Smith, Liz Tashjian, and seminar attendees at the 2000 European Finance Association meetings, University of Utah, and the Australian Graduate School of Management for their useful comments.

[1] According to the *Wall Street Journal* (1999), 91 percent of actively managed stock funds generated lower returns than the S&P 500 index over the 10 years ending in December 1998, and 84 percent trailed the Wilshire 5000 over the same period.

ex post performance of timing strategies. Again, they find no evidence of timing ability.

In the studies mentioned thus far, observations of mutual fund returns are recorded monthly or annually. As discussed by Goetzmann, Ingersoll, and Ivković ((2000), hereafter referred to as GII), a monthly frequency might fail to capture the contribution of a manager's timing activities to fund returns, because decisions regarding market exposure are likely made more frequently than monthly for most funds.[2] We have daily observations of mutual fund returns. This allows us to directly overcome the problem investigated by Goetzmann et al. To determine whether observation frequency matters, we generate daily and monthly data under the null of no timing ability and under various alternatives and, at both observation frequencies, test the size and power of standard timing regressions. Both daily and monthly tests falsely reject the null at about the right rate for a given significance level. In all cases, however, the tests using daily data are more powerful than the monthly tests.

We analyze a set of mutual fund returns at both the daily and monthly frequencies to determine whether the use of daily data changes inference regarding managerial ability. The daily tests result in a larger number of significant estimates of timing ability, both positive and negative. Jagannathan and Korajczyk ((1986) hereafter referred to as JK) show that standard timing tests spuriously reject the null hypothesis of no ability if fund returns are more or less option-like than the market proxy. To control for this, we create a synthetic matched sample of funds that mimics the holdings of the actual funds but that have no timing ability by construction. Using one model of market timing and monthly data, 11.9 percent of the funds exhibit significantly more timing ability than the corresponding synthetic fund. Using daily data, 34.2 percent of the funds exhibit significantly more ability. A second timing model generates qualitatively similar inference. These results indicate that a substantial number of the funds in our sample possess significant timing ability.

This paper makes two main contributions to the mutual fund performance literature. First, we demonstrate that daily data provide different inferences than monthly data regarding timing ability.[3] Second, we provide evidence that the timing results cannot be explained simply as a spurious statistical phenomenon. In summary, our results motivate the use of daily data in future tests of mutual fund performance, and suggest that more fund managers possess market timing ability than previously documented.

The rest of the paper is organized as follows. Section I discusses the tests of timing ability used in the study. Section II describes the data. Section III examines the size and power of timing tests. Section IV presents the empirical analysis. Section V offers concluding remarks.

[2] As pointed out by the referee, standard tests of stock selection, which use the intercept of factor regressions to measure stock-picking ability, are theoretically robust to observation frequency because the estimate is more a function of sample length rather than observation frequency.

[3] Chance and Hemler (1999) document a similar result in a study of 30 professional market timers; however, their sample does not provide direct evidence regarding the performance of mutual fund managers.

I. Tests of Market Timing Ability

Market timing refers to the dynamic allocation of capital among broad classes of investments, often restricted to equities and short-term government debt. The successful market timer increases the portfolio weight on equities prior to a rise in the market, and decreases the weight on equities prior to a fall in the market.[4] This section discusses the models we use to test for market timing ability.

Treynor and Mazuy (1966) use the following regression to test for market timing:

$$r_{p,t} = \alpha_p + \beta_p r_{m,t} + \gamma_p r_{m,t}^2 + \varepsilon_{p,t}, \tag{1}$$

where $r_{p,t}$ is the excess return on a portfolio at time t, $r_{m,t}$ is the excess return on the market, and γ_p measures timing ability. If a mutual fund manager increases (decreases) the portfolio's market exposure prior to a market increase (decrease) then the portfolio's return will be a convex function of the market's return, and γ_p will be positive.

Henriksson and Merton (1981) develop a different test of market timing. In their model, the mutual fund manager allocates capital between cash and equities based on forecasts of the future market return, as before, except now the manager decides between a small number of market exposure levels. We test a model with two target betas via the following regression:

$$r_{p,t} = \alpha_p + \beta_p r_{m,t} + \gamma_p r_{m,t}^* + \varepsilon_{p,t}, \tag{2}$$

where

$$r_{m,t}^* = I\{r_{m,t} > 0\} r_{m,t} \tag{3}$$

and $I\{r_{m,t} > 0\}$ is an indicator function that equals one if $r_{m,t}$ is positive and zero otherwise. The magnitude of γ_p in equation (2) measures the difference between the target betas, and is positive for a manager that successfully times the market. We use both timing models to measure timing ability in our sample of mutual funds.

[4] A mutual fund manager's ability to shift a fund's allocation is constrained to varying degrees by the investment objectives of the fund, as established in the fund's "Statement of Additional Information." A manager constrained to holding equities might then time the market by adjusting the correlation between a portfolio's return and the market return as the market rises and falls. In addition, market timing activity may be hindered by restrictions on the use of leverage and derivatives placed on mutual funds by the Securities and Exchange Commission's Investment Company Act of 1940. Hedge fund managers are not constrained by these sorts of limits; hence we may expect more evidence of market timing and other dynamic strategies among hedge fund managers than mutual fund managers, as indicated by the results of Fung and Hsieh (1997).

Grinblatt and Titman (1994) show that tests of performance are quite sensitive to the chosen benchmark. For this reason, we run four-factor analogs of equations (1) and (2) in which the three additional factors are the Fama and French (1993) size and book-to-market factors and Carhart's (1997) momentum factor. The additional factors have been shown to capture the major anomalies of Sharpe's (1964) single-factor CAPM, and are included so as not to reward managers for simply exploiting these anomalies.[5] The three additional factors appear only as linear terms; we do not estimate "factor timing" except for the market factor. We express the four-factor TM regression as

$$r_{p,t} = \alpha_p + \sum_{i=1}^{4} \beta_{p,i} r_{i,t} + \gamma_p r_{m,t}^2 + \varepsilon_{p,t}, \tag{4}$$

and the four-factor HM regression as

$$r_{p,t} = \alpha_p + \sum_{i=1}^{4} \beta_{p,i} r_{i,t} + \gamma_p r_{m,t}^* + \varepsilon_{p,t}. \tag{5}$$

We estimate parameters of the two models using both daily and monthly data to determine whether observation frequency affects inference regarding market timing ability.

Scholes and Williams (1977) point out that when estimating the parameters of a factor model of daily stock returns, infrequent trading can result in biased estimates of variance, serial correlation, and contemporaneous correlation between assets. This holds for portfolios of infrequently traded assets as well, because the variance of a portfolio is largely determined by the average covariance of the individual assets in the portfolio. When using daily data, we use Dimson's (1979) correction and include lagged values of the factors as additional independent variables in the regressions to accommodate infrequent trading.

II. Data

We study daily returns of 230 mutual funds. The sample, taken from Busse (1999), is constructed as follows. A list of all domestic equity funds with a "common stock" investment policy and a "maximum capital gains," "growth," or "growth and income" investment objective and more than $15 million in total net assets is created from the December, 1984, version of Wiesenberger's (1985) *Mutual Funds Panorama.* Sector (e.g., technology or health care), balanced, and index funds are not included, nor are funds that changed into one of these types of funds in subsequent years during the sample period.

[5] Another approach to address the issue of benchmark efficiency is to use stochastic discount factors, as in Chen and Knez (1996), Dahlquist and Söderlind (1999), and Farnsworth et al. (1999).

Daily per share net asset values and dividends from January 2, 1985, through December 29, 1995, are taken from Interactive Data Corp., which acquires its net asset value data from the National Association of Security Dealers. *Moody's Dividend Record: Annual Cumulative Issue* (Moody's Investors Service, Inc. (1985–1995)) and *Standard & Poor's Annual Dividend Record* (Standard and Poor's Corporation (1985—1995)) are used to verify the dividends and dividend dates and to determine split dates. The net asset values and dividends are combined to form a daily return series for each fund as follows:

$$R_{p,t} = \frac{NAV_{p,t} + D_{p,t}}{NAV_{p,t-1}} - 1, \tag{6}$$

where $NAV_{p,t}$ is the net asset value of fund p on day t, and $D_{p,t}$ are the ex-div dividends of fund p on day t. Of the 244 funds in the December 1984 version of *Panorama* that meet the specified criteria, 230 funds are tracked through the end of the sample period or until merger or liquidation and are included in the sample. The returns of 14 funds could not be reconciled with Morningstar's monthly returns, and these funds are not included. This sample does not suffer from survivorship bias of the sort discussed in Brown et al. (1992) and Brown and Goetzmann (1995), wherein only funds in existence at the end of the sample period are included. However, funds that come into existence at some point between the end of 1984 and the end of the sample period are not included.

To determine whether daily data generates different inferences than monthly data, monthly returns are constructed from the daily returns as follows. Suppose there are N trading days in a particular month and let T denote the first day of the month. The monthly return R^M based on daily returns R^D is

$$R^M = \prod_{t=T}^{T+N-1} (1 + R_t^D) - 1. \tag{7}$$

Panel A of Table I lists summary statistics of the fund return distributions. We test the hypothesis that fund returns are normally distributed using the Jarque-Bera (1980) statistic, which is distributed χ_2^2 under the null. For the daily data, only one of the funds fails to reject normality at the one percent level. The average test statistic is 342,958, whereas a value of 9.21 or higher rejects the null. For the monthly data, the average test statistic is 217 and only four funds fail to reject normality.[6] Evidence of nonnormality in our mutual fund sample is relevant because of the Jagannathan and Korajczyk (1986) suggestion that option-like payoffs can generate spurious evidence of market timing. We will return to this issue when interpreting the results of our timing tests.

[6] These results should come as no surprise, since the nonnormality of stock returns is well established and has spurred the study of alternative distributional assumptions as well as the development of stochastic volatility models of returns.

Table I

Summary Statistics

Listed are average summary statistics of the 230 mutual funds in our sample and the market index. The sample period is January 2, 1985, to December 29, 1995, a total of 2,780 trading days or 132 trading months. The mean (μ) and standard deviation (σ) are sample estimates. Skewness (S) is computed as

$$S = \frac{1}{\sigma^3 T} \sum_{t=1}^{T} (R_t - \mu)^3$$

and excess kurtosis (K) is computed as

$$K = \frac{1}{\sigma^4 T} \sum_{t=1}^{T} (R_t - \mu)^4 - 3.$$

The Jarque-Bera (JB) test for normality is distributed χ_2^2 under the null and is given by

$$JB = \frac{T}{6}\left[S^2 + \frac{K^2}{4}\right]$$

Panel A: Daily Statistics

	μ	σ	S	K	JB Test
Mutual funds					
Daily	0.056%	0.898%	−2.504	48.580	342,958.079
Monthly	1.223%	4.756%	−1.074	5.454	217.104
Market proxy					
Daily	0.060%	0.846%	−3.408	71.419	596,217.106
Monthly	1.289%	4.202%	−1.345	6.847	297.661

Panel B: Annual Statistics

Year	# Funds	μ	σ
1985	221	28.9%	9.6%
1986	226	14.2%	12.7%
1987	224	6.1%	27.4%
1988	226	14.6%	13.2%
1989	219	26.3%	11.0%
1990	216	−4.8%	14.7%
1991	212	38.6%	14.0%
1992	204	8.3%	11.1%
1993	199	13.7%	10.1%
1994	197	−1.9%	10.7%
1995	194	32.4%	9.7%

Panel A of Table I also lists summary statistics for our market proxy, the CRSP value-weighted index including NYSE, AMEX, and Nasdaq stocks. The market index rejects normality at the daily and monthly frequency. Furthermore, the index exhibits higher excess kurtosis and larger negative skewness than the average of the mutual funds. The negative skewness is probably due to the crash of 1987 and other smaller crashes in the sample. Again, the relative degree of nonnormality in the mutual funds and the market index may explain some of the market timing results, as we discuss in Section IV. Panel B of Table I shows the number of funds in the sample each year, as well as the average fund mean return and standard deviation of return. Note that the sample includes years of high and low returns, as well as a range of standard deviations, suggesting that the sample is rich enough to capture market timing activity.

In an effort to control for possible spurious results, we create for each fund in the sample a synthetic fund that matches fund characteristics but has no timing ability by construction. The synthetic funds are created as in Busse (1999). For each fund in the sample, we determine the fund's exposure to eight asset classes: the six intersections of the two equally weighted size and the three equally weighted book-to-market indices, the equally weighted momentum index, and the equally weighted contrarian index. If we express fund p's return on date t as

$$r_{p,t} = \sum_{i=1}^{8} b_{p,i} r_{i,t} + \varepsilon_{p,t} \tag{8}$$

where $r_{i,t}$ is the return on asset class i on date t, then the bs are selected by minimizing the variance of ε_p, subject to a nonnegativity constraint on the bs. Given these weights on the asset classes, a synthetic fund is constructed by randomly selecting 100 stocks chosen from the different asset classes in proportions to match the fund's vector of bs. The stocks are initially equally weighted. We replace stocks by other stocks in the same asset class at random, with an average holding period of one year. When a stock is replaced, weights are reset to equal weight. Between replacements, weights evolve according to a buy and hold strategy. This procedure is similar in spirit to the way Daniel et al. (1997) create characteristic-based benchmarks in order to test for managerial ability, except that Daniel et al. use their funds' quarterly holdings rather than a quadratic program to determine asset class exposures.

We construct monthly and daily versions of the size and book-to-market factors similar to the monthly factors of Fama and French (1993). We construct monthly and daily versions of the momentum factor similar to the monthly factor of Carhart (1997). The Appendix explains how we construct the daily versions of these factors.

We use the 90-day U.S. Treasury bill index on Datastream (code *TBILL90*) to estimate the return on the riskless asset.

In addition, to compare our daily tests to the GII monthly tests, we reconstruct Goetzmann et al.'s (2000) monthly factor that proxies for the monthly payoffs of a successful market timer. The value of the monthly factor is computed each month as

$$P_{m,t} = \left(\prod_{\tau=1}^{N} \max\{1 + R_{m,\tau}, 1 + R_{f,\tau}\} \right) - 1 - R_{m,t}, \tag{9}$$

where there are N days in month t, $R_{m,\tau}$ is the market return on day τ, and $R_{f,\tau}$ is the riskless return. This factor is then used in the following regression using monthly returns to capture correlation between a fund's monthly return and the monthly value of daily timing:

$$r_{p,t} = \alpha_p + \sum_{i=1}^{4} \beta_{p,i} r_{i,t} + \gamma_p P_{m,t} + \varepsilon_{p,t}, \tag{10}$$

where the four factors are those used in the TM and HM models. This regression corresponds to the three-factor model that Goetzmann et al. (2000) label the "adjusted-FF3" test.

III. Statistical Properties of Tests of Timing Ability

In this section, we generate mutual fund returns under the null hypothesis of no timing ability and under two alternatives in order to gauge the size and power of the timing tests. We find that the tests are substantially more powerful when applied to daily data rather than monthly. This provides motivation for the next section, in which we estimate the timing ability of actual mutual funds using daily data.

We examine the size of the tests by generating fund returns under the null hypothesis of no timing ability. First, we estimate parameters of a four-factor model of stock returns applied to the daily set of actual mutual fund returns using OLS and save the residuals. The four-factor model is similar to the timing models described in Section I except without the timing terms:

$$r_{p,t} = \alpha_p + \sum_{i=1}^{4} \beta_{p,i} r_{i,t} + \varepsilon_{p,t}. \tag{11}$$

Second, we generate 1,000 sets of daily returns for each fund under the null hypothesis of no timing ability. On each date the fund is in existence, we randomly draw one of the fund's residuals with replacement and add it to the fund's fitted return from the estimated non-timing model. We generate monthly data by compounding the daily data. Third, we estimate parameters of the two timing models on the simulated daily and monthly data and assess individual fund timing significance at the five percent level using standard OLS t-statistics. The resampling procedure ensures that residuals from this last step are free of serial correlation and heteroskedasticity, and that the generated returns do not reflect timing strategies.

Panel A of Table II shows the results of the size tests. The table lists the fraction of funds generated under the null hypothesis that result in positive and negative timing coefficients and the fraction that result in significantly positive and significantly negative timing coefficients. The size of the daily tests appears correct for both models, with half of the coefficients positive and about five percent significant. The significant coefficients are equally split between positive and negative. The monthly tests appear somewhat biased, however, with between 55.6 percent and 58.3 percent positive, and with about twice as many positive significant coefficients as negative significant coefficients.

We examine the power of the tests by generating fund returns under two alternative hypotheses of either TM or HM timing ability. Our goal is to demonstrate that increasing the frequency with which returns are recorded can increase power.

To generate returns under the TM alternative, we construct a time series of fund betas as

$$\beta_{p,t:t+T} = \beta_p + \gamma \bar{r}_{m,t:t+T}, \tag{12}$$

where $\bar{r}_{m,t:t+T}$ is the mean daily excess market return from day t until day $t + T$, and $t{:}t + T$ represents the manager's timing interval (one day, two days, one week, two weeks, or one month). β_p is the fund's beta from the non-timing model of equation (11). We substitute the beta from equation (12) into the non-timing model and add a randomly sampled residual (from the non-timing model regression) to generate a fund return under the TM alternative. We generate returns by setting γ equal to 5, 7.5, 10, 15, and 20. These values result in mild to aggressive trading behavior. Consider, for example, a monthly timing interval. A large monthly return for the market is on the order of five percent. The lowest level of γ we consider, five, corresponds in this case to an increase in fund β of 0.25; the highest level of γ we consider, 20, corresponds to an increase in fund β of 1.0.

In the HM timing simulations, we take the market beta of a perfect timer to be

$$\beta_{p,t:t+T} = I\{\bar{r}_{m,t:t+T} > 0\}\beta_p. \tag{13}$$

We substitute the beta from equation (13) into the non-timing model and add a randomly sampled residual (from the non-timing model regression) to generate a fund return under the HM alternative. We also run simulations for imperfect timing ability by choosing beta according to equation (13) for a fraction, $0.6 < p < 0.9$, of the timing decisions. For the remaining $1 - p$ of the timing decisions, we choose beta incorrectly,

$$\beta_{p,t:t+T} = I\{\bar{r}_{m,t:t+T} \leq 0\}\beta_p. \tag{14}$$

We run three timing models (TM, HM, and GII; all four-factor) on the daily data and monthly data generated under the TM and HM alternatives and assess individual fund timing significance at the five percent level using standard t-statistics.

Panels B and C of Table II show the results of our power tests when data are generated with a weekly timing frequency. Panel B shows the results for data generated under the TM alternative. The tests result in a positive timing coefficient in most cases, but the daily tests result in significantly pos-

Table II

Size and Power Analysis

Panel A summarizes timing coefficients from the four-factor Treynor and Mazuy (1966; TM), Henriksson and Merton (1981; HM), and Goetzmann et al. (2000; GII) timing models applied to fund returns generated under the null hypothesis of no timing ability. Returns are generated by randomly reordering residuals from a nontiming four-factor model of returns. Listed for each model is the fraction of simulated funds with positive (significantly positive) or negative (significantly negative) timing coefficients. Significance is at the five percent level (two-tailed). The timing models are all of the form

$$r_{p,t} = \alpha_p + \sum_{i=1}^{4} \beta_{p,i} r_{i,t} + \gamma_p f(r_{m,t}) + \varepsilon_{p,t},$$

where r is excess return, $f(r_{m,t}) = r_{m,t}^2$ for TM, $f(r_{m,t}) = \mathrm{I}\{r_{m,t} > 0\} r_{m,t}$ for HM, and $f(r_{m,t})$ is the value of a monthly timing factor constructed from daily index returns for GII. The GII factor is computed as

$$\left(\prod_{\tau=1}^{N} \max\{1 + R_{m,\tau}, 1 + R_{f,\tau}\} \right) - 1 - R_{m,t},$$

where there are N days in the month and R is return. Panel B shows the results when returns are generated under the TM alternative. Returns are generated by constructing a time series of fund betas that are a linear function of the market return, $\beta_{p,t:t+T} = \beta_p + \gamma \bar{r}_{m,t:t+T}$, where $t{:}t + T$ is a weekly timing interval. Panel C shows the results when returns are generated under the HM alternative. Returns are generated by constructing a time series of betas of either zero or the nontiming regression beta, depending on the sign of the contemporaneous market return. P denotes the fraction of observations for which the timing decision is made correctly.

Panel A: Size

	Daily		Monthly	
	Positive	Negative	Positive	Negative
TM	0.499 (0.022)	0.501 (0.023)	0.583 (0.046)	0.417 (0.018)
HM	0.499 (0.024)	0.501 (0.025)	0.556 (0.037)	0.444 (0.020)
GII	—	—	0.573 (0.045)	0.427 (0.020)

Panel B: Treynor-Mazuy

	Daily, γ					Monthly, γ				
	5.0	7.5	10.0	15	20	5.0	7.5	10.0	15	20
TM	0.970	0.978	0.973	0.972	0.971	0.932	0.980	0.988	0.994	0.998
	(0.919)	(0.947)	(0.951)	(0.954)	(0.958)	(0.342)	(0.613)	(0.794)	(0.932)	(0.975)
HM	0.983	0.990	0.996	0.998	0.999	0.885	0.958	0.979	0.991	0.996
	(0.847)	(0.942)	(0.966)	(0.976)	(0.982)	(0.257)	(0.464)	(0.663)	(0.864)	(0.946)
GII	—	—	—	—	—	0.921	0.980	0.987	0.997	0.998
						(0.346)	(0.612)	(0.778)	(0.927)	(0.975)

Panel C: Henriksson-Merton

	Daily, p					Monthly, p				
	0.6	0.7	0.8	0.9	1.0	0.6	0.7	0.8	0.9	1.0
TM	0.600	0.701	0.789	0.906	0.985	0.589	0.732	0.837	0.941	0.999
	(0.592)	(0.695)	(0.781)	(0.896)	(0.976)	(0.359)	(0.481)	(0.634)	(0.832)	(0.976)
HM	0.659	0.835	0.969	0.998	1.000	0.660	0.822	0.909	0.978	0.999
	(0.614)	(0.774)	(0.911)	(0.992)	(1.000)	(0.298)	(0.464)	(0.666)	(0.874)	(0.981)
GII	—	—	—	—	—	0.700	(0.373)	0.851	(0.580)	0.941
						(0.766)	0.987	(0.926)	1.000	(0.998)

itive timing coefficients much more often than the monthly tests, for all but the most extreme market timer. For example, with a timing coefficient of $\gamma = 5$, the daily tests generate significantly positive coefficients about 92 percent of the time using the TM model and about 85 percent of the time using the HM model. The monthly tests result in a significantly positive coefficient in only 34 percent of the funds generated under the TM alternative using the TM model, and only 26 percent of the time using the HM model. The additional factor suggested by GII, which is designed to improve the HM test, increases the frequency of significantly positive coefficients, but only to about 35 percent of the time. As the magnitude of the timing ability increases, the monthly tests improve. Panel C shows the results for data generated under the HM alternative. A similar pattern emerges: the daily tests result in significant timing coefficients much more often than the monthly tests.

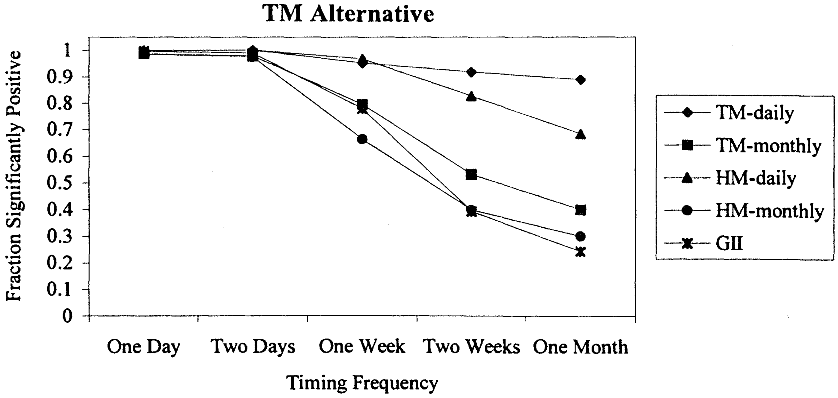
Figure 1 displays these power results graphically for several frequencies of timing activity. When we generate data under the TM specification, the daily data do not provide an advantage over the monthly data when market timing occurs daily or every two days. However, as the timing frequency decreases, the relative power of the daily tests increases. This is likely the result of higher precision from an increased number of observations. When we generate data under the HM specification, the correctly specified daily test dominates the monthly tests at all timing frequencies. The GII test outperforms the incorrectly specified TM daily test at high frequency timing, but the TM daily test outperforms the GII test when timing occurs at a weekly interval or less.

In summary, the power tests show that daily tests correctly reject the null of no timing ability more often than monthly tests. We turn next to an analysis of the mutual fund sample to measure actual timing ability.

IV. Empirical Analysis

A. Bootstrap Standard Errors

Assessing the significance of the actual funds' timing regression coefficients is complicated by the possibility of misspecification of the timing function or of timing strategies that change over time. For example, if a fund manager times the market according to the TM model, but we measure timing ability using the HM specification, we will likely induce temporary serial correlation in the residuals while the strategy is being executed. Furthermore, there is evidence that fund managers execute timing strategies dynamically. For example, Brown, Harlow, and Starks (1996) suggest that fund managers may change investing strategies over the calendar year depending on year-to-date performance, in an effort to game compensation schemes. Also, Busse (1999) provides evidence that fund managers time exposure to the market to coincide with low levels of market volatility. Misspecifying the timing function may cause violations of regression assumptions in unknown and possibly time-varying ways, so that standard corrections



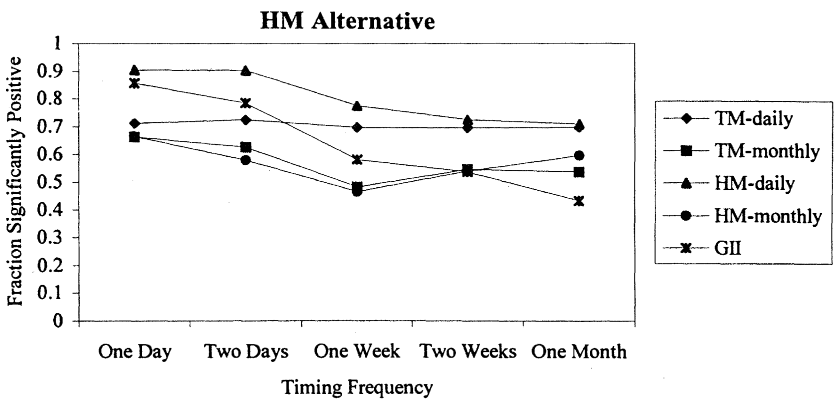


Figure 1. Power analysis. The figure shows the results from running four-factor Treynor and Mazuy (1966; TM), Henriksson and Merton (1981; HM), and Goetzmann et al. (2000; GII) timing models on 1,000 sets of fund returns generated under the alternative hypothesis of timing ability. The timing models are all of the form

$$r_{p,t} = \alpha_p + \sum_{i=1}^{4} \beta_{p,i} r_{i,t} + \gamma_p f(r_{m,t}) + \varepsilon_{p,t},$$

where r is excess return, $f(r_{m,t}) = r_{m,t}^2$ for TM, $f(r_{m,t}) = \mathrm{I}\{r_{m,t} > 0\} r_{m,t}$ for HM, and $f(r_{m,t})$ is the value of a monthly timing factor constructed from daily index returns for GII. The GII factor is computed as

$$\left(\prod_{\tau=1}^{N} \max\{1 + R_{m,\tau}, 1 + R_{f,\tau}\} \right) - 1 - R_{m,t},$$

where there are N days in the month and R is return. Returns are generated under the TM alternative by constructing a time series of fund betas that are a linear function of the market's excess return, $\beta_{p,t:t+T} = \beta_p + 10\bar{r}_{m,t:t+T}$, where $t:t + T$ is the timing interval. Returns are generated under the HM alternative by constructing a time series of betas of either zero or a nontiming regression beta depending on the sign of the contemporaneous market excess return. The beta is selected correctly 70 percent of the time to model imperfect timing ability. The figures show the fraction of simulated funds with significant positive timing coefficients. Significance is at the five percent level (two-tailed).

for heteroskedasticity and serial correlation may not fully capture the effect of these violations on the standard errors of regression coefficients.

To overcome this statistical problem, we construct bootstrap standard errors for the timing coefficients following the procedure described by Freedman and Peters (1984). There are three steps in this procedure. First, for each fund, we estimate parameters of the TM and HM timing models using daily and monthly data over the 1985 to 1995 sample period. Second, we generate bootstrap fund returns fund-by-fund as follows. For each date that a fund is in existence, we randomly choose with replacement one of the fund's residuals and add it to that date's fitted return from the original timing regressions. We repeat the process 1,000 times, resulting in 1,000 sets of bootstrap returns for each fund. The third step is to estimate parameters of the timing models on each set of bootstrap data. For each fund, then, we have 1,000 timing coefficients for both timing models and both observation frequencies. The standard error of each fund's 1,000 timing coefficients is the bootstrap standard error of the original timing coefficient, which we use to compute empirical t-statistics of the form

$$t = \frac{\gamma_{p,original}}{\sigma(\gamma_{p,bootstrap})}. \tag{15}$$

We assess significance at the five percent level and so compare the empirical t-statistic to ± 1.96, the critical value under the assumption of normality.[7]

B. Empirical Results

Table III lists the fraction of funds that have positive and negative timing coefficients and the number of funds that have significantly positive and negative timing coefficients. Displayed are the results from daily and monthly data. Panel A shows the results for the mutual fund sample. In all cases, the fraction of funds with significant timing ability is higher when daily data are used instead of monthly. For the TM model, for example, 40.8 percent of the funds generate significantly positive coefficients and 28.1 percent produce significantly negative coefficients using daily data. The corresponding frequencies using monthly returns are 33.5 percent and 5.3 percent. The HM model gives similar results. The daily data's higher rejection rate is consistent with our power analysis and suggests that there is a wide dispersion of ability over the sample of funds.

A conservative interpretation of the results requires the consideration of two potential sources of spurious timing coefficients. One possible source of spurious timing ability is the cash-flow hypothesis described in Warther

[7] We also assess significance by sorting the bootstrap distribution of timing coefficients by size, and comparing the magnitude of the actual timing coefficient to the 25th and 975th bootstrap timing coefficients. This avoids the distributional assumption. The results are almost identical.

Table III

Bootstrap Analysis of Market Timing Coefficients

Listed are the fraction, mean timing coefficient, and mean intercept of 230 mutual funds that exhibit positive/negative (+/-) and significant positive/significant negative (++/−−) market timing abilities. The sample period is January 2, 1985, to December 29, 1995, a total of 2,780 trading days or 132 trading months. The intercepts are converted to annualized percentages. Timing ability is measured using the four-factor Treynor and Mazuy (1966; TM) and Henriksson and Merton (1981; HM) timing models. The timing models are of the form

$$r_{p,t} = \alpha_p + \sum_{i=1}^{4} \beta_{p,i} r_{i,t} + \gamma_p f(r_{m,t}) + \varepsilon_{p,t},$$

where r is excess return, $f(r_{m,t}) = r_{m,t}^2$ for TM, and $f(r_{m,t}) = \mathrm{I}\{r_{m,t} > 0\} r_{m,t}$ for HM. Significance is at the five percent level (two-tailed) and is based on bootstrap standard errors. Panel A shows the results from the mutual fund sample, and Panel B shows the results from the synthetic control sample. A synthetic fund is constructed under the null hypothesis of no timing ability for each fund in the sample by selecting stocks to match the fund's style and randomly replacing the stocks by others in the same asset class. Panel C shows the fraction of funds for which the difference between the fund's timing coefficient and the timing coefficient of the corresponding synthetic fund is positive/negative (+/-) and significantly positive/significantly negative (++/−−).

	Monthly		Daily		Monthly		Daily	
	+	−	+	−	++	−−	++	−−
Panel A: Mutual Fund Sample								
Fraction								
TM	0.736	0.264	0.561	0.439	0.335	0.053	0.408	0.281
HM	0.771	0.229	0.592	0.408	0.256	0.035	0.382	0.184
Timing coefficient								
TM	0.716	−0.274	0.899	−0.440	1.126	−0.785	1.174	−0.587
HM	0.185	−0.088	0.086	−0.044	0.336	−0.189	0.123	−0.069
Intercept								
TM	−0.248	0.544	−0.830	0.196	−0.161	1.146	−0.842	0.153
HM	−2.499	0.894	−4.726	2.185	−3.480	3.007	−6.659	3.920
Panel B: Synthetic Control Sample								
Fraction								
TM	0.925	0.075	0.684	0.316	0.317	0.004	0.360	0.105
HM	0.925	0.075	0.702	0.298	0.313	0.004	0.360	0.066
Timing coefficient								
TM	0.385	−0.180	0.253	−0.194	0.601	−1.626	0.378	−0.344
HM	0.120	−0.060	0.032	−0.029	0.185	−0.202	0.046	−0.073
Intercept								
TM	−1.006	−0.098	−0.938	−0.123	−1.529	2.995	−1.139	0.069
HM	−2.324	0.457	−2.513	1.423	−3.541	4.091	−3.472	4.165
Panel C: Difference in Timing Coefficients								
Fraction								
TM	0.480	0.520	0.482	0.518	0.119	0.088	0.342	0.333
HM	0.471	0.529	0.496	0.504	0.062	0.084	0.281	0.259

(1995), Ferson and Warther (1996), and Edelen (1999). The hypothesis suggests we might bias timing coefficients downwards, even to negative levels, because when market returns are high, investors increase subscriptions to mutual funds, resulting in a temporarily larger cash position and a lower fund beta. Warther (1995) finds a strong relation between a fund's cash inflows and its portfolio weight on cash. Ferson and Warther (1996) show directly that changes in conditional fund betas are negatively related to changes in fund cash flows. Edelen (1999) shows that monthly fund cash flows can completely explain monthly estimates of negative timing ability. We do not have daily cash flow data for our sample of funds, so we cannot control for this possible effect, and leave this task for future research.

Note, though, that the cash-flow explanation is asymmetric in the sense that it can bias timing coefficients downwards but not upwards. For the HM specification, the timing coefficient is estimated using returns that occur when the market's excess return is positive. If the cash position of the fund increases during these times, the timing coefficient will be biased downwards reflecting the decrease in beta. For the TM specification, the timing coefficient is estimated in times of both market rises, when subscriptions to the fund likely increase, and market declines, when we might expect fund redemptions to increase. In the former case, the timing coefficient will be biased downwards following the same argument as in the HM specification. In the latter, we might expect an increase in beta, because cash reserves become depleted, which serves to bias the timing coefficient downwards again. The reason for this is that in the TM specification, the timing coefficient weights the squared market return. In times of negative market excess returns, we expect fund returns to be lower than they would be without the redemptions; hence this forces the timing coefficient to be lower than it would be otherwise. Because the cash flow explanation postulates that timing coefficients will be biased downwards, our results may underestimate the true ability of fund managers in the sample.

The other possible source of spurious timing is provided by Jagannathan and Korajczyk (1986), who argue that spurious timing ability can be generated when portfolios hold stocks with payoffs that are more or less option-like than the market proxy. In particular, if the average stock in a mutual fund is more option-like than the average stock in the market proxy, a timing regression will result in a positive timing coefficient and a negative intercept, which is usually interpreted as measuring the stock-selection ability of the fund manager. Recall from Table I that the mutual funds exhibit less negative skewness than the market proxy on average. We might expect states in which mutual fund returns and market returns are both negative, due to their correlation, and in which the market return is more negative than the mutual fund returns, due to its larger negative skewness. These states would generate a positive timing coefficient even in the absence of market timing activity.

In Panel A of Table III, there does appear to be an inverse relation between the timing coefficients and intercepts in the timing regressions as predicted by Jagannathan and Korajczyk (1986). In all cases, the average intercept for the funds with negative timing coefficients is much higher than the corresponding average for funds with positive timing coefficients. Kon (1983) and Henriksson (1984) also document a negative correlation between regression intercepts and timing coefficients. Both find that most mutual funds in their respective samples exhibit positive intercepts and negative timing coefficients, the reverse of what we find, perhaps due to differences in our sample periods. To test the relation more formally, we regress intercepts on timing coefficients cross-sectionally for each timing model. For the daily data, the slope is negative and significant for both timing models, indicating that estimates of stock selection and market timing are significantly negatively related. This result suggests that some of the positive timing coefficients in our sample could be spurious.

In an effort to control for the JK source of spurious timing ability, we run the timing tests on a sample of synthetic funds that match the actual funds' characteristics but have no timing ability by construction, as described in Section II. If the synthetic funds exhibit timing ability at the same frequency and magnitude as the actual funds, then the estimated timing coefficients are likely spurious rather than evidence of ability.

Panel B of Table III shows the results of our timing tests when applied to the synthetic funds. Using both monthly and daily data, the synthetic funds exhibit more significant timing coefficients than expected under the null of no timing activity. For the TM model, for example, 36.0 percent of the synthetic funds have a significantly positive timing coefficient and 10.5 percent have a significantly negative coefficient using daily data. This suggests that some of the timing evidence for the actual funds is spurious, likely the result of the JK phenomenon. However, note that when using daily data, the actual funds reject the null more frequently than the synthetic funds. Furthermore, the magnitude of the average significantly positive daily timing coefficients using the actual fund returns is roughly three times larger than the average using the synthetic fund returns. For the HM model, for example, the average positive significant timing coefficient is 0.123 for the actual funds and 0.046 for the synthetic funds. This indicates that although the synthetic funds exhibit significant timing coefficients, their magnitude is likely insufficient to fully explain the timing coefficients of the actual funds.

We investigate the relation between the timing ability of actual funds and their synthetic counterparts more formally by computing the difference between their timing coefficients. We assess significance by constructing a standard error for the difference from the bootstrap standard errors of the timing coefficients as follows:

$$\sigma(\text{difference}) = \sqrt{\sigma^2(\gamma_{\text{actual}}) + \sigma^2(\gamma_{\text{synthetic}})}. \tag{16}$$

Panel C of Table III shows the results. Using monthly returns and the TM model, 11.9 percent of the funds have timing coefficients that are significantly larger than their synthetic counterparts, and 8.8 percent of the funds' timing coefficients are significantly smaller than the synthetic funds'. Using daily data and the TM model, however, 34.2 percent of the funds have timing coefficients that are significantly larger than the synthetic ones, and 33.3 percent of the funds have coefficients that are significantly smaller. The results for the HM model are qualitatively similar. The cash-flow hypothesis can explain the significant negative differences. To the extent that the synthetic funds control for spurious rejections of the null, the significantly positive differences suggest that a substantial percentage of the funds have true timing ability.

V. Conclusions

In this paper, we demonstrate that using daily rather than monthly data changes inference regarding the market timing ability of mutual fund managers. We first document that standard regression-based tests have more power to detect significant timing activity when daily data are used. We then estimate timing coefficients for a sample of mutual funds and find that daily returns increase the number of significant estimates of timing ability. To test whether this result is spurious, we construct a set of synthetic funds that match the characteristics of the actual funds but have no timing ability. Using one model of market timing and monthly data, 11.9 percent of the funds exhibit significantly more timing ability than the corresponding synthetic fund. Using daily data, 34.2 percent of the funds exhibit significantly more ability. A second model of market timing generates qualitatively similar inference. These results indicate that the measured timing ability cannot be explained as a spurious statistical phenomenon.

Observation frequency matters when judging fund performance, suggesting that future research in mutual fund performance may generate more precise estimates and sharper inference if daily data are used rather than data collected at a lower frequency.

Appendix

We construct the size and book-to-market indices following the procedure used by Fama and French (1993), except with daily returns instead of monthly.

We sort all firms listed on both CRSP and COMPUSTAT and classified as having ordinary common shares (on CRSP) according to market capitalization at the end of June each year beginning in June of 1983. As in Fama and French (1993), to mitigate the problems associated with COMPUSTAT's practice of back filling data, firms must exist on COMPUSTAT for two years before we use them. We take market capitalization to be the number of shares as of the end of June (per CRSP) multiplied by the end of June CRSP share price. We also sort these same firms according to their end of calendar year book-to-market ratio, where we take book value as the COMPUSTAT book value of shareholders' equity plus balance sheet deferred taxes and investment tax credit, minus the book value of preferred stock. We take the book value of preferred stock to be the redemption, liquidation, or par value (in that order) on COMPUSTAT.

We use NYSE breakpoints to divide firms into two groups, big (B) and small (S), where the big group includes all firms (NYSE, AMEX, and Nasdaq) greater than or equal to the median market capitalization of NYSE firms. We also use NYSE breakpoints to divide all firms into three groups, high book-to-market (H), medium book-to-market (M), and low book-to-market (L), depending on each firm's book-to-market relative to the 70th and 30th percentiles of NYSE firms.

Combining the two market capitalization groups with the three book-to-market groups results in six groups of firms: one that includes big firms with high book-to-market ratios, one with big firms and medium book-to-market ratios, one with big firms and low book-to-market ratios, and an analogous set of three groups of small capitalization firms. We compute a return index for each of the six groups by weighting the returns by market capitalization.

We form the size index by taking the difference between an equal weighted combination of the three small market capitalization indices and the three big market capitalization indices. We form the book-to-market index by taking the difference between an equal weighted combination of the two high book-to-market indices and the two low book-to-market indices.

We construct the momentum index similar to that of Carhart (1997), except at a daily frequency. For each month t, we rank all firms on CRSP (NYSE, AMEX, and Nasdaq) classified as having ordinary common shares with returns for a month t-12 to t-2 evaluation period by total return from t-12 to t-2. We take the momentum index for month t as the difference between the equal weighted month t return index of the 30 percent of firms with the highest returns during the evaluation period and the equal weighted index of the 30 percent of firms with the lowest returns during the evaluation period. We reallocate firms to the 30 percent highest returns and 30 percent lowest returns groupings monthly.

REFERENCES

Brown, Keith C., W. V. Harlow, and Laura T. Starks, 1996, Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry, *Journal of Finance* 51, 85–110.

Brown, Stephen J., and William N. Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679–698.

Brown, Stephen J., William N. Goetzmann, Roger G. Ibbotson, and Stephen A. Ross, 1992, Survivorship bias in performance studies, *Review of Financial Studies* 5, 553–580.

Busse, Jeffrey A., 1999, Volatility timing in mutual funds: Evidence from daily returns, *Review of Financial Studies* 12, 1009–1041.

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

Chance, Don M., and Michael L. Hemler, 1999, The performance of professional market timers: Daily evidence from executed strategies, Working paper, Virginia Tech.

Chen, Zhiwu, and Peter J. Knez, 1996, Portfolio performance measurement: Theory and applications, *Review of Financial Studies* 9, 511–555.

Dahlquist, Magnus, and Paul Söderlind, 1999, Evaluating portfolio performance with stochastic discount factors, *Journal of Business* 72, 347–383.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.

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

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

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.

Farnsworth, Heber, Wayne E. Ferson, David Jackson, and Steven Todd, 1999, Performance evaluation with stochastic discount factors, Working paper, University of Washington.

Ferson, Wayne E., and Vincent A. Warther, 1996, Evaluating fund performance in a dynamic market, *Financial Analysts Journal* 52, 20–28.

Freedman, David A., and Stephen C. Peters, 1984, Bootstrapping an econometric model: Some empirical results, *Journal of Business and Economic Statistics* 2, 150–158.

Fung, William, and David A. Hsieh, 1997, Empirical characteristics of dynamic trading strategies: The case of hedge funds, *Review of Financial Studies* 10, 275–302.

Goetzmann, William N., Jonathan Ingersoll Jr., and Zoran Ivković, 2000, Monthly measurement of daily timers, *Journal of Financial and Quantitative Analysis* 35, 257–290.

Graham, John R., and Campbell R. Harvey, 1996, Market timing ability and volatility implied in investment newsletters' asset allocation recommendations, *Journal of Financial Economics* 42, 397–421.

Grinblatt, Mark, and Sheridan Titman, 1994, A study of monthly mutual fund returns and performance evaluation techniques, *Journal of Financial and Quantitative Analysis* 29, 419–444.

Henriksson, Roy D., 1984, Market timing and mutual fund performance: An empirical investigation, *Journal of Business* 57, 73–96.

Henriksson, Roy D., and Robert C. Merton, 1981, On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills, *Journal of Business* 54, 513–533.

Investment Company Institute, 2000, *Mutual Fund Fact Book* (Investment Company Institute, Washington DC).

Jagannathan, Ravi, and Robert A. Korajczyk, 1986, Assessing the market timing performance of managed portfolios, *Journal of Business* 59, 217–235.

Jarque, Carlos M., and Anil K. Bera, 1980, Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals, *Economics Letters* 6, 255–259.

Kon, Stanley J., 1983, The market-timing performance of mutual fund managers, *Journal of Business* 56, 323–347.

Moody's Investors Service, Inc., 1985–1995, *Moody's Dividend Record: Annual Cumulative Issue* (Moody's Investors Service Inc., New York, NY).

Scholes, Myron, and Joseph T. Williams, 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics* 5, 309–327.

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

Standard and Poor's Corporation, 1985–1995, *Standard and Poor's Annual Dividend Record* (Standard and Poor's Corporation, New York, NY).

Treynor, Jack L., and Kay Mazuy, 1966, Can mutual funds outguess the market? *Harvard Business Review* 44, 131–136.

Wall Street Journal, 1999, Debunking some mutual-fund myths, March 16, C1.

Warther, Vincent A., 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209–235.

Weisenberger, 1985, *Mutual Funds Panorama* (Warren, Gorham, & Lamont, Boston, MA).