An Examination of Mutual Fund Timing Ability Using Monthly Holdings Data

EDWIN J. ELTON[1], MARTIN J. GRUBER[1], and CHRISTOPHER R. BLAKE[2]
[1]*New York University,* [2]*Fordham University*

Abstract

In this paper, the authors use monthly holdings to study timing ability. These data differ from holdings data used in previous studies in that the authors' data have a higher frequency and include a full range of securities, not just traded equities. Using a one-index model, the authors find, as do two recent studies, that management appears to have positive and statistically significant timing ability. When a multiindex model is used, the authors show that timing decisions do not result in an increase in performance, whether timing is measured using conditional or unconditional sensitivities. The authors show that sector rotation decisions with respect to high-tech stocks are a major contribution to negative timing.

JEL Classification: G11, G12

1. Introduction

While a large body of literature exists on whether active portfolio managers add value, the vast majority of this literature has concentrated on stock selection.[1] In its simplest terms, this literature examines how much better a manager does compared to holding a passive portfolio of securities with the same risk characteristics (sensitivities to one or more indexes). The bulk of the literature on performance measurement ignores whether managers can time the market as a whole or time across subsets of the market, such as industries. By doing so, that literature assumes that either timing does not exist or, if it does exist, it will not distort the measurement of an analyst's ability to contribute to performance through stock selection.

A number of articles have shown that the existence of timing on the part of management can lead to incorrect inference about the ability of managers to pick stocks whether evaluation is based on either single-index or multiple-index tests of performance.[2] Because of this possibility, and because of the importance of timing ability as an issue, some papers have been written that explore the ability of managers to

---

[1] See, for example, Elton, Gruber, and Blake (1996), Gruber (1996), Daniel *et al.* (1997), Carhart (1997), Zheng (1999), and references therein.
[2] See, for example, Dybvig and Ross (1985) and Elton *et al.* (2010b) for discussions on how timing can lead to incorrect conclusions about management performance.

successfully time the market. This literature started with the work of Treynor and Mazuy (1966), who explore whether there was a nonlinear relationship between the market beta with the market and the return on the market. That work was followed by Henriksson and Merton (1981), who look at changes in betas as a reaction to discrete changes in the market return relative to the Treasury bill rate. Other studies follow, using more sophisticated measures of the return-generating process, to examine how time series sensitivities of mutual fund returns vary with market and factor returns.[3]

The potential problem with almost all these studies is that they assume management implements timing in a specific way. (For example, Henriksson and Merton (1981) assume a different but constant beta according to whether the market return is lower or higher than the risk-free rate.) If management chooses to time in a more complex manner, these measures may not detect it. To overcome the estimation problem caused by the assumption of a specific form of timing, two recent studies (Jiang, Yao, and Yu, 2007, and Kaplan and Sensoy, 2008) estimated portfolio betas using portfolio holdings and security betas. They find, using a single-index model, that mutual funds have significant timing ability. These findings are opposite to what prior studies have found. The purpose of this paper is to see if these findings hold up when holdings data and security betas are used to measure timing in a multiindex model.

We collect data on the actual holdings of mutual funds at monthly intervals. This allows us to construct the beta or betas on a portfolio at the beginning of any month using fund holdings. As explained in more detail later, this is done by using 3 years of weekly data to estimate the betas on each stock in a portfolio and then using the actual percentage invested in each security to come up with a portfolio beta at a point in time. We refer to the portfolio betas constructed this way as "bottom-up" betas.

This approach differs from that which has been taken in the literature with respect to timing measures with the exception of the two articles that found positive timing ability: Jiang, Yao, and Yu (2007) (hereafter JY&Y) and Kaplan and Sensoy (2008) (hereafter K&S). While our paper follows in the spirit of these articles, we believe that our methodology is an improvement over theirs in several ways. First, both articles investigate only the effect of changing betas in a single-index model. In addition to the one-index model, we examine a two-index model that recognizes bonds as a separate vehicle for timing, the Fama–French model (with the addition of a bond index), both with unconditional and conditional betas, and a model that examines the impact of changing allocation across industries.[4] As we show, the use

---

[3] See, for example, Bollen and Busse (2001), Chance and Hemler (2001), Comer (2006), Ferson and Schadt (1996), and Daniel *et al.* (1997).

[4] We report results for the two-index model. The results, while similar to the results for the one-index model, do vary for certain funds that hold bonds. We also examined the Fama–French model with the Carhart (1997) momentum factor added. The conclusions reached are similar to the ones reported without the momentum factor.

of a more complete model leads to conclusions that are different from those reached when the single-index model is used. The reason for this is that when managers change their exposure to the market, they often do so as a result of shifting their exposure to small stocks or higher growth stocks. When the effect on performance of these shifts is taken into account, timing results change. In particular, the positive timing ability identified with the use of a one- or two-index model becomes negative timing ability. Second, we examine monthly data rather than quarterly holdings data as used in prior studies. The use of quarterly data misses 18.5% of the round-trip trades made by the average fund manager.[5] Third, we account for timing using a full set of holdings including bonds, nontraded equity, preferred stock, other mutual funds, options, and futures. The database used by JY&Y, but not K&S, forced them to assume that all securities except traded equity have the same impact on timing. In particular, JY&Y assume the beta on the market of all securities that are not traded equity is zero. Thus, nontraded equity, bonds, futures, options, preferred stock, and mutual funds are all treated as identical instruments, each having a beta on the market of zero. As we show, using the full set of securities rather than only traded equity results in very different timing results. We follow this with a section that examines management's ability to time the selection of industries. We find that reallocating investments across industries decreases performance and that most of this decrease in value is explained by mistiming the tech bubble.

In the first part of this paper, we examine the ability of monthly holdings data to detect timing ability using unconditional betas. We show that inferences about timing ability differ according to whether a single-index or multiindex model is used and the single-index model does not result in an accurate measure of timing ability. Next, we examine measures of timing ability that are conditional on publicly available data. Following the general methodology of Ferson and Schadt (1996) (hereafter F&S), we find that employing a set of variables that measures public information explains a large part of the action management takes with respect to systematic risk and changes the conclusions about timing ability. This is direct evidence that mutual fund management reacts to macrovariables that have been shown to predict return and also provides additional evidence that using holdings data to measure management behavior is important. The use of conditional timing measures results in estimates that are closer to zero than unconditional measures.

This paper is divided into eight sections. The next section after the introduction discusses our sample. That section is followed by a section discussing our methodology. In the Section 4, we discuss timing results using unconditional betas. That

---

[5] See Elton *et al.* (2010a) for details on the amount of trades missed using different frequencies of holding data. While we describe the Thomson database as containing quarterly holdings data, in many cases, the actual holdings are reported at much linger intervals. For our sample, more than 16% of the time Thomson reported holdings at semiannual or longer intervals.

section is followed by a section discussing the reasons for differences in results between alternative models of the return-generating process, a section discussing timing across industries, and a section discussing the effects of using conditional betas. The final section presents our conclusions.

2. Sample

Data on the monthly holdings of individual mutual funds were obtained from Morningstar. Morningstar supplied us with all its holdings data for all of the domestic (USA) stock mutual funds that it followed anytime during the period 1994–2004. The only holding Morningstar does not report is that of any security that represents less than 0.006% of a portfolio and, in early years in our sample, holdings beyond the largest 199 holdings in any portfolio. This has virtually no effect on our sample since the sum of the weights almost always equals 1 and, in the few cases where it was less than 1, the differences are minute.[6]

Most previous studies of holdings data use the Thomson database as the source of holdings data (K&S is an exception). The Morningstar holdings data are much more complete. Unlike Thomson data, Morningstar data include not only holdings of traded equity but also holdings of bonds, options, futures, preferred stock, other mutual funds, nontraded equity, and cash. Studies of mutual fund behavior from the Thomson database ignore changes across asset categories such as the bond/stock mix and imply that the only risk parameters that matter are those estimated from traded equity securities. While this can affect any study of performance, the drawback of these missing securities is potentially severe when measuring timing.[7]

From the Morningstar data, we select all domestic equity funds, except index and specialty funds, that report holdings for at least 8 months in any calendar year, did not miss two or more consecutive months, and existed for at least 2 years. These are funds that report monthly holdings most of the time but occasionally miss a month.

---

[6] While Morningstar in early years reports only the largest 199 holdings in a fund, this does not affect our results since most of the funds that held more than 199 securities were index funds, and we eliminate index funds from our sample since they do not attempt timing.

[7] Like other studies, the funds in our sample have a high average concentration (over 90%) in common equity. This is used by others to justify using a database that has no information on assets other than traded equity. However, average figures hide the large differences across funds and over time. Twenty-five of the funds in our sample use futures and options, with the future positions being as much as 40% of total assets. Over 20% of the funds vary the proportion in equity by more than 20%, and they differ in the investments other than equity that are used when equity is changed. The funds that have variation in the percent in equity over time or use assets that can substantially affect sensitivities are precisely the ones that are likely to be timing. Thus, in a study examining timing, it is important to have information on all assets the fund holds.

Only 4.6% of the fund months in our sample do not have data, on average 57% of the fund years have complete monthly data, and 96% of the fund years are not missing more than 2 months. Less than 1% of the funds have only 8 months of monthly data in any 1 year.[8] Our sample size is 318 funds and 18,903 fund months.

An important issue is whether restricting our sample to funds that predominantly reported monthly holdings data or requiring at least 2 years of monthly data introduces a bias. This is examined in some detail in Elton *et al.* (2010a) and Elton, Gruber, and Blake (2011), but a summary is useful.

There are two possible sources of bias. First, funds that voluntarily provide monthly holdings data may be different from those that do not. Second, even if funds that provide monthly holdings are no different from those that do not, requiring at least two consecutive years of holdings data may bias the results. When we require 2 years of monthly holdings data, we are excluding funds that merged and excluding funds that reported monthly holdings data in 1 year but did not report monthly data in the subsequent year. Each of these potential sources of bias will now be examined.

The first question is whether the characteristics of funds that voluntarily report holdings monthly are different from the general population. In Table I, we report some key characteristics of our sample of funds compared to the population of funds in Center for Research in Sector Price (CRSP), which fall into each of the four categories of stock funds that we examine. The principal difference between our sample and the average fund in the CRSP is the average total net asset (TNA) value. Our sample's TNA is on average smaller. This is caused by the presence of a few gigantic funds in CRSP that are not in our sample. If we compare the median size, the CRSP funds have a median TNA less than 2.5% higher than our sample's median TNA. Turnover and expense ratios are also somewhat smaller for our sample.[9] The distribution of objectives of funds is almost identical between our sample and the CRSP funds.

For our study, it is the possibility of differences in performance and merger activity that needs to be carefully examined. For each fund in our sample, we randomly select funds with the same investment objective that did not report monthly holdings data. Using the Fama–French model, the difference in average alpha between our sample and the matching sample was 3 basis points, which is not statistically significant at any meaningful level. We also check merger activity. There were slightly fewer mergers in the funds that do not report monthly, but in any economic or statistical sense, there was no difference.

---

[8] The data included monthly holdings data for only a very small number of funds before 1998, so we started our sample in that year. In 1998, 2.5% of the common stock funds reporting holdings to Morningstar reported these holdings for every month in that year. By 2004, the percentage had grown to 18%.

[9] These differences are similar in magnitude to those found by Ge and Zheng (2006), who examined whether funds that report quarterly are different from funds that report annually.

Table I. Summary statistics of fund characteristics in 2002

This table shows the value of certain attributes of the funds in our sample as well as the value of those same attributes for funds in the CRSP database that have the same objectives as our sample funds.

Statistic	Sample	All funds
Number of funds	318	2,582
TNA (millions)	$386	$591
Turnover	0.82	1.09
Expense ratio	1.25	1.30
Aggressive growth, %	5.35	7.62
Growth, %	61.95	60.48
Growth and income, %	23.90	22.14
Balanced, %	8.81	9.76

Another bias could arise by requiring 2 years of monthly data if funds stopped reporting monthly holdings data because their performance changed or they realize that they were not performing as well as the funds that continued to report monthly data. For the funds that met our criteria in the first year but not in the second, 4 switched to quarterly reporting and 24 merged in the second year. Using standard time series regressions and the Fama–French model, we find that the four funds that switched to quarterly reporting perform no worse than the funds that continue to report holdings on a monthly basis. The 24 funds that meet reporting requirements in 1 year and merge in the second are on average poor performing funds. Examining our measures over the periods these funds exist shows timing results very slightly below what we report. Thus, our measures are very slightly biased upward. The evidence suggests that our sample does not differ in any meaningful way from the population of funds.

3. Methodology

There are two ways a manager can affect performance beyond security selection. First, the manager can vary the sensitivity of the portfolio to general factors such as the market or the Fama–French factors. This can be done by switching among securities of the same type but with different sensitivities to the factors or by changing allocation to different types of securities (e.g., stocks to bonds or preferred stocks). Second, the manager can vary the industry exposure, overweighting in industries that are forecasted to outperform others (usually called "sector rotation"). Clearly, these are interrelated. For example, managers engaged in sector rotation are likely to affect sensitivity to systematic market factors. However, it is useful to examine these separately and then to examine the joint implications of the two types of results.

3.1. TIMING AS FACTOR EXPOSURE

One way that management can make timing decisions is to change the sensitivity of the portfolio to a set of aggregate factors that affect returns. Because we have monthly holdings data, we can measure the sensitivity of a portfolio to any influence in successive months over the time period of interest.

A general model for mutual fund returns can be described by a multifactor model of the form:

$$R_{Pt} - R_{Ft} = \alpha_P + \sum_{j=1}^{J} \beta_{Pjt} I_{jt} + \epsilon_{Pt}, \tag{1}$$

where R_{Pt}, the return on mutual fund P in month t; R_{Ft}, the return on the 30-day T-bill in month t; I_{jt}, the return on factor j in month t (see below); β_{Pjt}, the sensitivity of fund P to factor j in month t; α_P, the risk-adjusted excess return on fund P; and ϵ_{Pt}, the residual return on portfolio P in month t.

Normally, the model is estimated by running a time series regression of the excess return on a fund against the excess return on a set of factors over time. However, this method suffers from the fact that if management is trying to engage in timing, the β_{Pjt} will vary over time. With holdings data, we can estimate the value of β_{Pjt} at a point in time by calculating the betas for each security in the portfolio and weighting the security betas by the percentage that security represents of the portfolio at that point in time.[10] The betas estimated in this manner are the unconditional betas. It has been shown that there are macrovariables that can predict returns, and it is argued that since the values of the macrovariables are known, management should not be given credit for changes in beta in response to those macrovariables. Thus, we will also estimate conditional betas. The exact method used in this estimation will be presented in the section on timing using conditional betas.

We now turn to the problem of choosing the factors in Equation (1). We first examine the simplest model used in the literature: the single-index model. However, since a number of funds in our sample have significant investments in bonds, we also use and emphasize a two-factor model containing an index of excess returns over the riskless rate for bonds and an excess-return index for stocks. The third model we use is a four-factor model consisting of the familiar Fama–French factors

---

[10] The betas or individual securities are estimated by running regressions on each security against the appropriate factor model using 3 years of weekly data ending in the month being estimated. There is clearly estimation error in the betas of individual securities. This estimation error tends to cancel out and becomes very small when we move to the portfolio level and examine measures over time. See Elton, Gruber, and Blake (2011) for a more detailed discussion and for estimates of the effect. The β_{Pjt} are exactly the same as would be obtained if one estimated them using a time series regression with fund returns if the weights remained unchanged over the estimation period.

with the excess return on a bond index added.[11] In Appendix A, we describe the details of estimating the models on different types of securities and the procedure we use for missing data.

How do we measure timing? Our timing measure is exactly parallel to the differential return measure used in measuring security selection ability. For each fund, we examine the differential return earned by varying beta over time rather than holding a constant beta equal to the overall average beta for that fund in our sample period.

For any model, the timing contribution of any variable j is measured by

$$\frac{I}{T} \sum_{t=1}^{T} \left[\beta_{Pjt} - \beta_{Pjt}^{*} \right] \times I_{jt+1}, \tag{2}$$

where β_{Pjt}^{*} is the target beta and T is the number of months of data available. When we use unconditional betas, the target beta is the average beta for the portfolio over the entire period for which we measure β_{Pjt}. I_{jt+1} is the excess return or differential return for factor j for the month following the period over which the beta is estimated. This intuitive measure of timing simply measures how well a manager did by varying the sensitivity of a fund to any particular factor compared to simply keeping the sensitivity at its target level. For any fund, this can be easily measured for each factor or for the aggregate of factors used in any of the models we explore.

This measure is very closely related to the measure utilized by Daniel *et al.* (1997). While we examine the current beta relative to the average beta, they use as a measure of differential exposure the difference in beta between the current beta and the beta 12 months ago. Each measure has some advantages. We use the average beta because, if the managers have a target beta, the mean is a good estimate of it, and deviation from a target beta is usually what we mean by timing.

In addition, as explained later, we use a conditional measure of the target beta. In this case, the deviations then become the difference between each month's estimated bottom-up beta and the target beta where the target beta is the expected value of beta adjusted for macrovariables.

3.2. CHANGES IN INDUSTRIES HELD

The availability of monthly holding data also allows us to look directly at whether changes in the allocations over time across industries improve performance. The

---

[11] We also added the Carhart momentum factor to this model. The conclusions are not substantially different, and where interesting are presented in the paper. All factors except for the bond index were provided by Ken French on a weekly basis. The bond index we use is the Lehman U.S. Government/ Credit index.

methodology directly follows that described in Section 3.1 above, but β_{Pjt} is replaced with X_{Pjt}, the fraction of the portfolio P in industry j at time t. The new measure for any industry is

$$\frac{I}{T}\sum_{t=1}^{T}\left[X_{Pjt}-\bar{X}_{Pjt}\right]\times I_{jt+1}, \tag{3}$$

where X_{Pjt} is the fraction of mutual fund P invested in industry j at time t, $\bar{X}_{Pjt}$ is the average amount invested in industry j by fund P, I_{jt+1} is the excess return on industry j at time $t+1$ the month following the reported holdings, and T is the number of months of data.

We divide equity holdings of the funds into five industry groups as designed by Ken French and available on his Web site.[12] Since we are interested in changes in stock allocation between industries, we normalize the industry weights at each point in time to add to one.

4. Evidence of Timing Unconditional Betas

Table II shows, for two versions of Equation (1), the average difference between the return earned on the factors using the funds' actual betas at the beginning of each month and the return they would have earned if they had held the sensitivities to the factors at their average values over the time period for which we have data. The average difference across funds is broken down into the average difference due to timing on each of the factors and the aggregate of these influences (called "overall"). Table II is computed over the 318 funds in our sample. The results for the one-index model are the same as those for the first index in the two-factor model. This comes about because the bond index and stock market index are virtually uncorrelated. Thus, in the interest of space, we only present results for the two-factor model. For the two-factor model, the average difference shows positive timing ability of approximately 5 basis points per month. This is similar to the results found by JY&Y. Examining the components of overall timing for the two-factor model shows that this extra return is almost entirely due to the timing of the stock market factor. Of the 318 funds, 233 showed positive timing ability. In order to examine the probability that the 5 basis points could have arisen by chance, we performed the bootstrap procedure described in Appendix B. The procedure is similar to the simulation procedure developed by Koswoski *et al.* (2006) (hereafter KTW&W) and the procedure employed by JY&Y. The purpose of the procedure is to examine

---

[12] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Similar results were obtained when we used the 17-industry classification designed by French.

Table II. Differential returns due to timing (average differences across 318 funds in %)

This table shows the differential return earned by funds through changing individual factor betas as well as the aggregate effect of these changes. A fund's factor- timing return is calculated as the fund's factor loading each month minus the target beta (the average factor loading over its entire sample period) times the leading monthly factor return. Overall is simply the sum of the individual factor timing returns. The two-factor model uses the Fama–French market factor (excess return over T-bill) and the excess return on the Lehman aggregate bond index. The four-factor model uses the three Fama–French factors (excess market, ''small-minus-big (SMB),'' and ''high-minus-low (HML)'' factors) and the excess return on the bond index.

	Mean	Median
Two factor		
Overall	0.0520	0.0740
Market	0.0517	0.0742
Bond	0.0003	0.0000
Four factor		
Overall	−0.1073	−0.0515
Market	−0.0247	−0.0130
Size (''SMB'')	−0.0572	−0.0221
Value/growth (''HML'')	−0.0261	−0.0213
Bond	0.0006	0.0000

statistical significance when it is likely that fund behavior is correlated. The simulation involves each month selecting at random a vector of actual factor returns and applying it to the actual differential betas that occurred in that month for each fund and then averaging over all months for each fund. Since the random assignment of a set of factor returns for each month is expected to produce a zero measure of timing, the 318 fund timing measures represent one possible set of outcomes when there is no timing. We repeat this 1,000 times to get 1,000 estimates of the timing measures when no timing exists in the data. This allows us to estimate the probability that any point on the distribution of actual values could have arisen by chance.

In Table III, we present the results of our simulation procedure. Note from Panel A that the probability of positive timing existing with the two-index model is extremely high. Let us explain the entries in the table. Consider the data under the entry 90%. For our 318-fund sample, the 32nd highest timing measure is the 90% cutoff value. To compute the associated probabilities, we take this value and compute the percentage of times across 1,000 simulations that a higher value occurs. For the 90th percentile, as shown in Table III, the simulation produced a higher value only 6% of the time. For the median and points on the distribution above the median, a *p* value is stated as the probability of getting a higher value than the associated cutoff value from our sample. For cutoff values below the median, a *p* value is stated as the probability of getting that value or lower. We follow KTW&W in also reporting the "significance" of the *t* values of the timing measures because, as they point out, *t* values have advantageous statistical properties.

The results from Panel A are clear. Most points of the distribution of actual values above the medium and the median itself are positive and significant at close to the 5% level. Whether we use raw timing measures or t values, the consistent pattern of p values for timing measures above the median indicate that the positive timing we found is unlikely to have arisen by chance.

When we examine the p values for points below the median, there is not much support for negative timing. Most p values are not close to any reasonable significant level. There are some funds that show negative timing, but the results could have arisen by chance. These are similar to the results found by JY&Y.

Our results in Table III use a different timing measure than JY&Y. They regress beta in period t on subsequent return (over 1, 3, 6, and 12 months). They use the slope of this regression as their measure of timing and found their strongest results using 3 months subsequent return. In order to see if the similarity in results held up when we use their measure, we repeat their analysis on our sample but use quarterly holdings, as they did, and use 3-month subsequent return. We find very similar results, a mean slope of 0.22, and a median of 0.27 compared to 0.35 and 0.31 for JY&Y. Table IV shows the simulation results for the JY&Y measure. The magnitude of the slopes is very similar to what they report (their Table III), but the level of significance is much higher. Almost all the cutoffs above the mean are significant, where they found significance only at the mean, median, and 75% cutoff rate.

As just discussed, these results are consistent in magnitude and statistical significance with those reported JY&Y, who examined timing ability for a different sample with a different methodology. However, using Thomson data at the most frequent interval available (usually quarterly) or Morningstar data monthly make a big difference in inferences about the timing behavior of individual funds. When we repeat our one-index analysis using Thomson data rather than Morningstar data, we find that 37% of the funds that were identified as good (or bad) timers using Morningstar monthly data were identified in the opposite group using all available Thomson data, quarterly or semiannual (when only semiannual was available). Of the seventy-one funds showing significant positive or negative timing ability (at the 5% level) using Thomson quarterly or semiannual data, only fifteen show significant positive or negative timing using monthly Morningstar data and four were significant in the opposite direction.

We find that the principal reason for the difference in performance of individual funds is that, as a fund changes its beta, this change was picked up by Morningstar by the end of the month, but it might not be picked up for 3 or 6 months using Thomson data.[13] This is illustrated in Figure 1, where we plot the data for one of the funds in our sample. The Thomson quarterly data indicate that this fund is a negative timer with a p value of -0.027, while Morningstar monthly data

---

[13] Recall that Thomson reports holdings at semiannual or longer intervals more than 16% of the time.

Table III. Statistical significance of timing measures

This table shows the timing measure and *t* value of the timing measure at various points on the distributions across the 318 sample funds and the probability they could have occurred by chance. For the median and all points above the median, the *p* value is the probability of a higher value occurring by chance. For points below the median, the *p* value is the probability of the value or lower occurring by chance. All probabilities are calculated using the simulation described in Appendix B.

	5%	10%	20%	40%	Mean	Median	60%	80%	90%	100%
Panel A: two-index model										
Timing (*p*)	−0.153 (0.27)	−0.099 (0.41)	−0.033 (0.66)	0.044 (0.90)	0.052 (0.13)	0.075 (0.06)	0.094 (0.06)	0.143 (0.06)	0.183 (0.06)	0.210 (0.07)
t (*p*)	−1.41 (0.41)	−1.12 (0.42)	−0.33 (0.71)	0.66 (0.92)	0.67 (0.11)	1.00 (0.06)	1.26 (0.04)	1.61 (0.06)	1.94 (0.06)	2.18 (0.06)
Panel B: four-index model										
Timing (*p*)	−0.508 (0.01)	−0.372 (0.03)	−0.218 (0.15)	−0.089 (0.29)	−0.107 (0.16)	−0.051 (0.67)	−0.021 (0.66)	0.037 (0.68)	0.096 (0.65)	0.138 (0.67)
t (*p*)	−2.14 (0.07)	−1.94 (0.06)	−1.50 (0.10)	−0.78 (0.20)	−0.52 (0.20)	−0.41 (0.73)	−0.18 (0.70)	0.30 (0.68)	0.80 (0.58)	1.18 (0.51)

Table IV. Significance using slope

This table shows the Jiang, Yao, and Yu (2007) timing measure and t value of the timing measure at various points on the distributions and the probability they could have occurred by chance. The timing measure is the slope of the regression of the market beta on market return in the subsequent 3 months. All timing measures are multiplied by 100. For the median and all points above the median, the p value is the probability of a higher value occurring by chance. For points below the median, the p value is the probability of the value or lower occurring by chance. All probabilities are calculated using the simulation described in Appendix B.

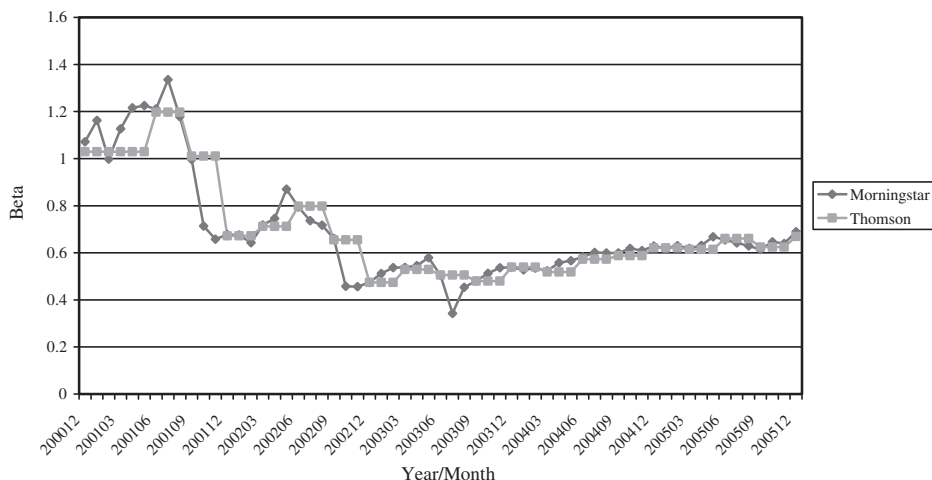|	5%	10%	20%	40%	Mean	Median	60%	80%	90%	100%
Timing (p)	−0.72 (0.36)	−0.49 (0.48)	−0.20 (0.75)	0.13 (0.96)	0.22 (0.94)	0.27 (0.02)	0.39 (0.01)	0.65 (0.02)	0.88 (0.03)	1.03 (0.06)
t (p)	−1.73 (0.39)	−1.26 (0.49)	−0.58 (0.72)	0.48 (0.97)	0.56 (0.95)	0.82 (0.01)	1.12 (0.01)	1.56 (0.04)	1.96 (0.07)	2.31 (0.09)

Figure 1. Monthly betas using Morningstar holdings data and quarterly betas using Thomson holdings data for one mutual fund.
This figure shows the portfolio beta for a fund calculated from portfolio holding weights and security betas. The portfolio holdings from Morningstar are available monthly and those from Thomson are available quarterly or semiannually. The portfolio betas should be almost identical at those quarterly or semiannual points in time when Thomson reports new holdings data. Differences at those points in time are due to differences between the two sources in reported holdings data or in reported total assets.

indicate that it is a significant positive timer with a *p* value of $+0.021$. While the principal difference in the results from the two databases was a delay in picking up a change in beta using quarterly or semiannual data rather than monthly data, there are other reasons for the difference. In several cases, the fact that Morningstar included preferred, debt, options, and futures, and Thomson did not, made a difference in the estimated beta. Finally, in some cases, there is a difference in some of the traded equity securities listed in the two databases. In cases where there were differences and holdings could be identified with forms filed with the Securities and Exchange Commission, Morningstar data more accurately matched actual holdings.

When we examine the four-factor model (Panel B in Table III), timing results are different. The difference in return due to timing the four factors is -11 basis points per month. In addition, 296 of the differentials are negative and 22 are positive. Examining the various factors shows that changing betas on the size factor is the major contributor to the negative timing.

Table III, Panel B, presents evidence of the probability that positive or negative timing measures, using the four-factor model, could have arisen by chance. It is clear from the table that there is no evidence that would support positive timing. However, while the median fund shows no significant evidence of negative timing ability, there

is statistically significant evidence that the lower tail of the distribution of mutual funds (lower 5% and 10%) exhibits negative timing ability that could not have arisen by chance. This is true both for the timing measure and for the t values of timing.

The results from the two- and four-factor models are completely different. The timing measure results combined with the simulation indicate that, if one uses a one- or two-factor model, mutual funds on average appear to exhibit positive timing ability at an economic and statistically significant level. When the four-factor model (the three Fama–French factors plus a bond market index) is used, there is no evidence of successful timing ability on the part of mutual funds on average, and there is evidence that 10% of the funds show significant negative timing ability.

5. Differences in Estimates of Market Timing

In this section, we will present evidence on why the four-factor model is a more appropriate measure of market timing than the two-factor model. Let us start by examining two extreme ways management might be attempting to make timing decisions.

In the simplest approach, managers might be only making timing decisions on the sensitivity (beta) of the portfolio with the market and inadvertently neglecting the impact of their decisions on the other common factors that affect return, such as the change in the value/growth characteristics of the portfolio. Whether or not we believe that these are equilibrium factors, there is ample evidence that over time there are differential returns on value and growth and small and large firms that affect fund returns. Thus, inadvertently or not, changing sensitivities to these factors affects fund returns. Furthermore, as we show below, the market sensitivity of a portfolio is highly correlated with sensitivity to one of the other factors (value growth) that affects return. Without management action to control the sensitivities to other factors, a change in the market beta will change the other factor sensitivity and examining only the change in market beta will not correctly measure the total impact on return of a change in the market beta.

The other extreme is to assume that management is concerned with the impact on fund returns of changes in the sensitivity to all four factors in the return-generating process. In this case, the overall four-factor timing measure is appropriate because it measures the impact of changes in all the sensitivities in the return-generating process on returns.

In either case, the correct measure of the impact of management timing decisions should be measured by the four-factor model not by the two-factor model.[14] There

---

[14] The results for the four- and five-factor models are similar. We emphasize the four-factor model because, while funds make decisions to change the growth or size posture to aid in timing, we know of no funds that change momentum exposure as a timing device.

is another possibility: the manager is rewarded only for timing relative to the market. In this case, the manager may be shrewd in ignoring additional factors.

We will now provide evidence that management's market timing choice has a direct effect on their estimated timing choice for other factors. While the additional variables in the Fama–French model were designed to minimize the correlation with the market, the high-minus-low book-to-market factor (value minus growth) still has considerable correlation (-0.59) with the market.[15]

To understand the impact of market timing with respect to the value minus growth factor, we orthogonalize the value minus growth return index to the market return index and reran the analysis. The overall timing measure is unchanged. However, when we orthogonalize the value–growth index to the market index, it forces any comovement between these two measures to be attributed to value minus growth. The timing attributed to the market is (and must be mathematically) the same as it is in the two-index model (0.052). However, the timing measure associated with growth goes from -0.0261 to -0.0801 or a change of -0.054.

The difference in the value growth factor of -0.054 when we orthogonalize explains why the market timing measure changes from $+0.052$ with the two-index model to a negative number with the four-index model. If we do not orthogonalize, the change in the value growth timing measure (of -0.054) is captured in the market timing measure, changing it from plus to minus. This explains 76% of the change. The remainder is due to correlation between the other variables and the market.

While this explains what is going on mathematically, what does this mean for management? When management makes timing decisions with respect to the market and ignores changes in other factors, they may be changing the sensitivity of the portfolio to other dimensions of risk (e.g., size or value–growth). Over the period of our study, if they made timing decisions based on the two-index model, they would have, on average, been inadvertently making bad decisions with respect to other factors, particularly value–growth. Thus, what appeared to be good timing decisions, looking only at the market factor was actually hurting overall timing performance.

6. Industry Timing

As discussed earlier, a manager can add value by correctly estimating factor returns and switching the exposure to the factor in anticipation of the change in the factor

---

[15] Examination of the previous two 10-year periods using weekly data produced correlations only slightly lower than this period (-0.47 and -0.53).

Table V. Timing by industry (in %)

This table shows the differential return earned by changing the exposure to various industries rather than maintaining a constant exposure to each industry. In particular, the return due to changing the exposure is each month's actual beta times next month's industry return, while the return due to maintaining exposure is the average exposure times next month's industry return. Industries are defined by the five industry classification of Ken French. "Overall" is computed each month as the sum of the five industry differential returns.

	Mean	Median	Top quartile	Bottom quartile
Overall	−0.0742	−0.0556	0.0199	−0.2014
1. Consumer	−0.0096	−0.0074	0.0171	−0.0383
2. Manufacturing	−0.0091	−0.0084	0.0274	−0.0470
3. High tech	−0.0476	−0.0349	0.0552	−0.1671
4. Health	−0.0018	−0.0005	0.0238	−0.0292
5. Other	−0.0062	−0.0051	0.0295	−0.0436

return. A manager could also potentially add value by switching exposure to industry categories. The availability of holdings data allows us to explore whether managers have the ability to add value through changing their exposure to industry categories.

We accepted as a definition of relevant industries Ken French's five industry grouping of firms. The advantage of this definition is not only that French provides a rational and clear definition of the factors but also provides a long history of return series calculated for each industry. Once again we measured the manager's ability to successfully engage in industry timing (sector rotation) as the difference between the actual exposure at the beginning of the month minus the average exposure over the history of the fund times the leading return on the industry over the following month. These monthly differential returns are accumulated over the full history of each fund. Table V provides the overall measure of timing ability along with the timing ability with respect to each industry.

The overall timing measure from industry timing, shown in Table V, is negative and highly significant, whether we judge the average value by the mean or the median.[16] The mean is 33% lower than the median, which is caused by the distribution being left skewed and including some extremely poor timers. The bulk of the poor timing comes from bad decisions on one industry: high tech. When we examine the mean, 64% of the negative overall timing due to industry choice is caused by changing investment in high-tech stocks, while if we examine the median, 63% is due to changing investment in high-tech stocks. Management again

---

[16] We examine statistical significance using the same simulation methodology used to construct Table III. At all points in the distribution, we again see too many negative extremes to arise by chance. We repeated the analysis using French's 17-industry classification with similar results.

seems to exhibit negative timing ability, and the bulk of this negative timing ability comes from one industry: high tech.

Earlier we found that timing on the value–growth factor was a major component of the negative overall timing on the Fama–French factors. It is possible that this was due in large part to the timing of mutual funds' investment in the high-tech sector. To examine this, we run a regression of the Fama–French high-minus-low (HML) book-to-market value-growth factor returns on the five French industry sector portfolio (*S*) returns. The regression results are as follows:

$$HML = 0.817 + 0.367S_1 + 0.290S_2 - 0.483S_3 + 0.236S_4 - 0.354S_5$$
$$t(2.46)(1.89)(2.27)(- 7.35)(3.27)(- 1.37)$$

with a coefficient of determination of 0.46.

The size and *t* value of the sensitivity to the high-tech portfolio (S_3) and the average value of the returns in the high-tech industry group suggest that timing decisions by funds in the high-tech industry strongly influenced the timing results from the four-factor model.[17]

To examine more directly the impact of decisions about high-tech stocks on the timing measures using the four-factor model, we reproduced timing measures for our sample of mutual funds excluding all stocks in the high-tech industry (Industry 3). Weights were recalculated to maintain full investment. The results are presented in Table VI along with the previous results from Table II. Note that the overall mistiming measured by the model is reduced by almost 50%, and it is no longer statistically significant, while the mean of the mistiming measure on the value–growth factor changes sign.[18] With high-tech stocks included, management showed negative timing ability with respect to the value–growth factor. If these stocks are excluded from the portfolios, management shows positive timing ability with respect to the value–growth factor.[19] Thus, mistiming of the tech stocks explains about half of the overall negative timing shown by the four-factor model and in particular the negative timing of the Fama–French value–growth factor.[20]

---

[17] We also ran regressions of the market factor and the size factor against the five industry factors. The market was significantly loaded on all the industries, and the size factor had no statistically significant coefficient with any of the industry factors.

[18] Again we tested this using the same simulation methodology, we used to construct Table III. The only point on the distribution that was close to being significant was the 95% cutoff, which was significant at the 8% level. All the other points were insignificant.

[19] Looking at the betas of the funds on the high-tech industry, it is clear the funds added high-tech stocks late in the boom and were late in getting out. Recall that our sample period coincides with the high-tech bubble.

[20] The poor performance in timing the high-tech sector might have been the result of management's attempt to attract new cash inflows by investing in a hot sector and getting out of the sector when interest cools.

Table VI. Timing with and without tech stocks measured by the four-factor model (in %)

This table shows the differential return earned by changing the factor loadings on the Fama–French three-factor model plus a bond index compared to holding the factor loadings at their average value. ''Overall'' is the sum of the individual differential returns. The results are computed including and excluding tech stocks. When tech stocks are excluded, portfolio weights are rescaled to one.

	Overall	Market	Size	Value/growth	Bond
Including tech stocks					
Mean	−0.1073	−0.0247	−0.0572	−0.0261	0.0006
Median	−0.0515	−0.0130	0.0221	−0.0213	0.0000
Excluding tech stocks					
Mean	−0.0562	−0.0367	−0.0436	0.0233	0.0008
Median	−0.0283	−0.0291	−0.0169	−0.0007	0.0000

However, mutual funds still show negative returns from timing, but the results are not statistically significant.

While we attribute this change in timing ability to the high-tech sector, it may be due to a wider bubble in stocks that accompanied the high-tech bubble. The change due to the fact that there was a general bubble in stocks can be found by repeating the analysis including high-tech stocks but leaving out the years 1999 and 2000 in our analysis. When we do this, we find a decrease in the negative timing measure similar to that found when high-tech stocks are excluded.

This analysis points out the advantages of employing holdings data. Timing performance can be decomposed to a level that allows the structure of timing mistakes (or accomplishments) to be understood. By combining multifactor analysis with industry analysis, the reason that funds appear to be good timers or bad timers can be better understood.

7. Conditional Betas and Timing

F&S explore the impact on mutual fund performance of conditioning betas on a set of predetermined time-varying variables representing public information. F&S find that that conditioning beta on a small set of variables changes many of the conclusions about the selection and timing ability of mutual fund managers. They study timing in the context of a single-factor model, where the parameters of the model are measured from a time series regression of fund returns on market returns using both unconditional betas and betas conditioned on a set of variables measuring public information.

In previous sections, we examined the use of monthly bottom-up betas to measure timing. If changes in these bottom-up betas really measure management action over time and F&S are right that management changes its action based on a set of public-information variables, then these bottom-up betas should be strongly related

to the F&S variables. We examine this hypothesis in this section. The section can be thought of as a joint test of the efficacy of bottom-up betas as a measure of management behavior and the efficacy of the F&S variables in explaining management behavior.

7.1 THE CONDITIONAL VARIABLES

We follow F&S in defining four variables to capture public information that might affect management's choice of beta.[21] The variables are as follows:

(1) The 1-month Treasury bill yield lagged 1 month. To measure this, we use the 30-day annualized Treasury bill yield from the CRSP risk-free rates file. This yield is the rate on the bill that matures closest to 30 days.
(2) The dividend yield of the CRSP value-weighted index of New York and American stock exchanges stocks lagged 1 month. This is derived by dividing the previous 12 months of dividends by the price level of this index.
(3) The term spread lagged 1 month. This is measured by the yield on a constant maturity 10-year Treasury bond minus the yield on a 3-month Treasury bill.
(4) The quality (credit) spread in the corporate bond market lagged 1 month. This is measured by the BAA-rated corporate bond yield less the AAA corporate bond yield.

We follow F&S in assuming that time-varying betas in the four-factor model are a linear function of the four conditioning variables discussed above. If we designate these conditioning variables as Z_1 to Z_4, then the conditional beta with respect to any beta for fund P is found from the following time series regression:

$$\beta_{Pjt} = C_{P0j} + \sum_{k=1}^{4} C_{Pkj}Z_{kt} + \epsilon_{Pjt}, \qquad (4)$$

where β_{Pjt} is the bottom-up beta for portfolio P with respect to factor j at time t (which does not incorporate conditional information), C_{Pkj} is the regression coefficient of the jth factor on conditioning variable k for portfolio P, Z_{kt} is the value of conditioning variable Z_k at time t, and ϵ_{Pjt} is the random error term of the bottom-up beta for portfolio P with respect to factor j at time t.

7.2 THE IMPACT OF CONDITIONING VARIABLES ON MANAGEMENT BEHAVIOR

In order to examine whether management was changing beta in reaction to public information, we regress the bottom-up betas with respect to each factor for

---

[21] F&S also use a January dummy but find that it has virtually no effect, so we do not include it here.

Table VII. Four-factor bottom-up betas versus F&S variables

Panel A presents the average coefficient of determination of the regression of each of the bottom-up four-factor model coefficients against the four F&S variables. The column labeled ''Significant improvement in fit'' shows how many times the decrease in unexplained variance is statistically significant at the 0.05 level when the F&S variables are included. Panel B presents the number of times each Fama–French beta is related to each of the F&S variables at the 0.05 level. Note that the number of funds for the bond variable (β_4) is different from those for the other variables because we only include funds that have bonds in their portfolios.

		Panel A			Panel B		
		Goodness of fit		Number of times significantly related to			
Bottom-up betas	Number of funds	Significant improvement in fit	Average adjusted R^2	T-bill	Divided/price	Term	Credit spread
β_1	318	296	0.42	94	99	69	190
β_2	318	308	0.50	86	100	51	265
β_3	318	307	0.56	102	137	63	261
β_4	206	148	0.25	37	45	32	93

each fund against the F&S conditioning variables. The results are presented in Table VII.

Panel A shows the average (across all funds) coefficient of determination (R^2) of the bottom-up betas for each of the three Fama–French factors (β_1, β_2, and β_3) and the bond factor (β_4) with the F&S variables. For each of the Fama–French betas and the bond beta, between 25% and 56% is explained by the F&S conditioning variables. This is strong direct evidence that the F&S variables matter on average in explaining how funds change their betas. For the 318 funds, the F&S variables significantly (at the 5% level) reduce the unexplained variance of the bottom-up market beta 296 times, the small-minus-large beta 308 times, and the value minus growth beta 307 times. Not all funds include bonds in the portfolios. For the 206 funds that include bonds, the bond betas were significantly related to the F&S variables 72% of the time.

Of the four F&S variables, the variable that is most often significant is credit spread. Credit spread is significantly related to the market and small-minus-large betas and the relationship is primarily negative, while for the value–growth beta and the bond beta, the relationship is primarily positive. Thus, when the credit spread widens, funds generally lower their exposure to the market, small firms, and growth stocks, while increasing their exposure to large stocks, value stocks, and bonds.

The second most important variable is dividend over price. An increase in this variable causes funds to increase their small- and growth-stock exposure, while lowering the exposure to large and value firms. In general, an increase in T-bill

rates leads to an increase in exposure to value stocks relative to growth stocks, while an increase in the term premium causes funds to move from value to growth stocks.

Not all the signs and significance are consistent with empirical evidence of what predicts higher market returns. Thus, the F&S variables capture a mixture of funds using public information and research findings to predict market returns and simply behavioral reaction to macrovariables. The justification for removing the effect of changing beta using the F&S variables from the time pattern of beta changes is to not give funds credit for the impact of using public information on their actions. Insofar as the variables simply capture funds' reaction to a change in an economic variable and their reaction is inconsistent with what evidence shows predicts returns, this relationship should not be removed from the time series of betas. Thus, both the conditional and unconditional timing measures give insight into funds' timing behavior.

7.3 TIMING USING CONDITIONAL BETAS

The purpose of this section is to examine whether management timing actions, separate from management's reaction to public information, add value. Throughout this section, we concentrate on the four-factor model using bottom-up betas. We measure timing using Equation (2) but define the target beta (β_{Pjt}^*) as $\beta_{Pjt}^* = \widehat{C}_{P0j} + \sum_{k=1}^{4} \widehat{C}_{Pkj} \times Z_{kt}$, where the hats indicate regression estimates on bottom-up betas from Equation (4).

Table VIII shows that, when using conditioning information with the four-factor model, the size of the overall timing measure, while still negative, is reduced from the unconditional measure shown in Table II. The overall timing measure, while much closer to zero, still indicates some negative timing ability, but the difference from zero is insignificant at any of the break points in the simulation.[22]

In the prior section, we showed that the principal reason for the negative timing measure was the funds' attempted timing of the tech bubble. When we repeat the analysis of the four-factor model in Table VIII eliminating tech stocks, we obtain an overall timing measure that is positive (0.0008), exceedingly small, and indistinguishable from zero at any point in the simulated distribution.

We find, as did F&S, that, when using conditioning variables, the evidence of perverse timing is greatly diminished. Furthermore, any perverse timing that remains is entirely due to the choices made in tech stocks during the period of the high-tech stock bubble. These results hold using a different methodology to

---

[22] Using conditional betas, the overall timing measure using the two-factor model becomes negative and is insignificantly different from zero. We judge significance for both the two-factor and the four-factor models by repeating the simulation procedure and points in the distribution used in Table III. None of the points along the distribution were significant at any reasonable level; most were close to 50% (pure chance).

Table VIII. Differential return due to timing with conditional betas

This table parallels Table I except that the target beta is defined as the conditional top-down beta for each period.

	Mean	Median
Two factor		
Overall	−0.0054	−0.0066
Market	−0.0052	−0.0058
Bonds	−0.0002	0.0000
Four factor		
Overall	−0.0287	−0.0174
Market	−0.0055	−0.0053
Size	−0.0040	−0.0069
Value/growth	−0.0271	−0.0400
Bonds	−0.0001	0.0000

measure timing as well as a different sample and different time period than those used by F&S.

7.4 ESTIMATES USING THE TIME SERIES OF RETURNS

The calculation of bottom-up betas is very time consuming compared to estimating betas from a fund's time series of returns because the sensitivity of each security in the portfolio to the factors as well as the sensitivity of options and futures need to be computed. Furthermore, data on holdings to compute bottom-up betas are often not available or data are not correctly identified in the database. (For example, CUSIP (Committee on Uniform Security Identification Procedures) numbers are sometimes missing or incorrect.) Thus, it would greatly facilitate the analysis if the F&S variables regressed on the time series of returns could capture the changing betas observed using bottom-up betas.

F&S showed that conditional betas explained more of the time series of fund returns than did unconditional betas for a sample of 67 funds. In this section, we will examine the same issue for 318 funds. We will then see how much of the variation of bottom-up betas is explained by top-down conditional betas. The average adjusted R^2 from regressing individual fund returns on the four-factor model was 0.85. When conditional top-down betas are used in the time series regression, the adjusted R^2 increases to 0.884. This increase is very similar to the increase found by F&S. The conditional betas decrease unexplained variance by about 23%. Of the 318 funds, the conditional beta increased the explanatory power at a statistically significant level (using a 5% cutoff rate) for 159 funds. This conditioning of betas to the F&S variables does improve their ability to explain returns.

How similar are the beta estimates using bottom-up betas and the conditional top-down betas? We examined this in two ways. First, we simply regressed for

each fund the bottom-up betas on the conditional top-down betas. Second, we looked at consistency in sign and significance between the coefficients of the bottom-up betas and the conditional top-down betas. When we regress in time series the bottom-up betas on the conditional top-down betas estimated from the time series of fund returns, we get R^2 on average across all funds ranging from 0.18 for the beta on the market to 0.14 for the beta on the small-minus-large factor. When the bottom-up betas were regressed directly on the F&S variables, the F&S variables captured about 50% of the variation in bottom-up betas. Thus, conditional variables estimated from a time series of returns captures about one-third of the variation in bottom-up betas that is explained by using the F&S variables directly. Another way to examine the relationship of the top-down conditional betas and the bottom-up betas is to examine whether they capture the same relationship. Slope coefficients of 1,724 were significant when bottom-up betas were regressed on the F&S variables, 328 of the coefficients on the F&S variables were significant when the F&S variables were included directly in the regression of fund returns on the factors, and 131 were significant in both sets of regressions. Of the 131 that were jointly significant, 83 had the same sign. Thus, although the conditional variables capture some of the variation in bottom-up betas and top-down betas, the relationships between top-down betas and bottom-up betas and the F&S variables are quite different. Thus, using the F&S variables in combination with top-down betas only captures some of the information contained in bottom-up betas.

8. Conclusions

In this paper, we use data on the monthly holdings for a set of mutual funds to study the timing ability of these funds. By examining monthly holdings, we are able to see how management changes the risk parameters and industry holdings in a fund and to examine how this contributes to timing.

Our study differs from previous studies in both the methodology used and in the accuracy of the data. Other studies that use holdings data principally employ a database that includes only data on the holdings of publicly traded stock. Our database contains holdings of funds in options, futures, other mutual funds, preferred stock, bonds, and nontraded equity. Many funds use these additional instruments to time and ignoring their presence can lead to erroneous conclusions about management timing decisions. Furthermore, the few studies that use holdings data have used quarterly data rather than monthly data, which give at best a much coarser measure of timing.

Our major results are based on the Fama–French three-factor model with the addition of a bond factor. A portfolio's "bottom-up" beta with respect to any given factor at a point in time is calculated by multiplying the factor beta of each security in a portfolio by the fraction that security represents of the portfolio and then summing

across all the securities held by the fund. In addition, we extend the work of Ferson and Schadt (1996) to calculate conditional betas based on observable macrovariables.

We find some evidence that timing decisions result in a decrease in performance when timing is measured using conditional or unconditional sensitivities. However, the results are only statistically significant for the 10% of the worst timers using unconditional sensitivities. When we use conditional sensitivities, there is slight evidence of negative timing though these results are not statistically significant.

We find that sector rotation decisions result in negative timing measures. Examining the results for individual sectors shows that the majority of the negative impact on returns from sector rotation comes about because of a fund changing exposure to high-tech stocks. The funds in our sample invested in high-tech stocks late in the bubble and continued to invest heavily after it broke. Choices made with respect to high-tech stocks were also a major reason for the negative timing results when the four-factor model was used. This occurred in large part because of the correlation of the value–growth factor with returns on high-tech stocks. When we removed the effect of high-tech stocks from our data, management timing decisions have a smaller negative impact on timing, and when we use conditional betas with the high-tech stocks removed, timing decisions are indistinguishable from zero.

We also explored timing using a one-factor (the Fama–French excess stock market factor) model and a two-factor (the Fama–French excess stock market factor and the excess return on a bond market index) model. These models showed positive timing. However, choices on market sensitivity also impacted sensitivity choices on other variables that affect return. When these impacts are taken into account by using a multifactor model, the average timing measure is negative.

Appendix A: Bottom-up Holdings–Based Estimations

Our sample allows us to estimate the mutual fund betas from holdings data as frequently as monthly. To do this at any point in time, we estimate a time series regression (Equation (1)) using 3 years of weekly past return data on each common stock or mutual fund held by the mutual fund under study. There are two problems. First, if less than 36 months of data are available, we use as much data as are available unless it is less than 12 months. If we have less than 12 months of data available, we set the beta for the stock equal to the average beta for all other stocks in the portfolio. On average, this had to be done for less than 1.4% of the securities in any portfolio. The second problem involves the estimation of Equation (1) for securities other than common stock and mutual funds.

For T-bills and bonds with less than 1 year to maturity, we set all betas to zero. For each of the following categories of investments: long-term bonds, preferred stocks, and convertibles, we used an index of that category and obtained estimated betas by running a regression of the category index against the appropriate model.

Each bond, convertible, or preferred stocks is assumed to have the same beta as the relevant index. Finally, for options and futures, we use the same beta as the underlying instrument adjusted for leverage. We use the Black–Scholes formula to adjust the betas for leverage.

The beta for any fund can be found at a point in time by weighting the beta on each security held in the fund at that time by the percent that the security represents of the fund's portfolio.

Appendix B

The classic tests of statistical significance of the mean are based on the variance of means across funds. However, this measure may well overstate statistical significance due to funds reacting to the same information in the same way. Since we know that funds in the same family, and to a lesser extent funds in different fund families, have a tendency to change betas in the same direction, we need to take this into consideration. Koswoski *et al.* (2006) developed a simulation method to correct for this. This method was later applied by Jiang, Yao, and Yu (2007) in their timing study.

It is easiest to describe how the simulation is performed for a single-factor model. For each fund, for each actual month, we compute the differential beta as the actual beta minus the average beta over its life. This yields a matrix of differential betas with 318 rows, one for each fund and one column for each date. Then, a factor return is randomly drawn from actual data with replacement for each month. The timing measure is then computed as the average of factor returns times the differential beta over the history of each fund. Since the factor returns are drawn randomly, the expected timing measure should be zero, and in fact, they are very close to zero across our simulations. Note also that we have preserved any cross-fund pattern of beta changes. The simulation is repeated 1,000 times. The *p* values are computed for various cutoffs of the actual distribution by examining the number of times we get that cutoff or higher in the random simulations. For example, consider 90% cutoff in Panel A of Table III. For the actual sample, the 90% cutoff had a value of 0.183. We then examination all 1,000 simulations and count the number of times the 90% cutoff was that value or higher.

The same procedure is used for multiindex models except that when a month is selected at random, the values of all the indexes for that month are used in the simulation. This preserves the interrelationship between indexes that exist over time.

References

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

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

Chance, D. and Hemler, M. (2001) The performance of professional market timers: daily evidence from executed strategies. *Journal of Financial Economics* **62**, 377–411.

Comer, G. (2006) Hybrid mutual funds and market timing performance, *Journal of Business* **79**, 771–797.

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

Dybvig, P. and Ross, S. (1985) Differential information and performance measurement using the security market line. *Journal of Finance* **40**, 383–399.

Elton, E. J., Gruber, M. J., and Blake, C. R. (1996) The persistence of risk-adjusted mutual fund performance, *Journal of Business* **69**, 133–157.

Elton, E. J., Gruber, M. J., and Blake, C. R. (2011) Holdings data, security returns and the selection of superior mutual funds, *Journal of Financial and Quantitative Analysis,* **46**, in press.

Elton, E. J., Gruber, M. J., Blake, C. R., , Krasny, Y., and Ozelge, S. (2010a) The effect of holdings data frequency on conclusions about mutual fund behavior, *Journal of Banking and Finance* **34**, 912–922.

Elton, E. J., Gruber, M. J., Brown, S., and Goetzmann, W. (2010b) Modern Portfolio Theory and Investment Analysis, 8th edition. John Wiley & Sons, Hoboken, New Jersey.

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

Ge, W. and Zheng, L. (2006) The frequency of mutual fund portfolio disclosure. Working paper, University of California Irvine.

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

Henriksson, R. and Merton, R. (1981) On market timing and investment performance, *Journal of Business* **54**, 513–533.

Jiang, G., Yao, T., and Yu, T. (2007) Do mutual funds time the market? Evidence from holdings data, *Journal of Financial Economics* **88**, 119–145.

Kaplan, S. N. and Sensoy, B. A. (2008) Do mutual funds time their benchmarks? Working paper, University of Chicago.

Koswoski, R., Timmermann, A., Wermers, R., and White, H. (2006) Can mutual fund "stars" really pick stocks? *New Evidence from a Bootstrap Analysis* **61**, 2551–2596.

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

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