How do We Capture Long-Horizon Factor Timing?

Measures and their Application to Hedge and Mutual Funds

Yuekun Liu[a] and Alexey Malakhov[b]

October 2023

Abstract

We consider capturing multi-factor timing based on macroeconomic trends over long horizons by active portfolio managers. Building on Jensen (1972), we introduce a new measure of long-horizon timing activity, D_ALPHA, and compare it with extant timing measures. We highlight the difficulty of capturing long-horizon multi-factor timing by examining the out-of-sample efficacy of proposed measures with simulated timing data. We then apply the measures with documented efficacy in simulations to hedge and mutual fund data. Out of all considered measures, only D_ALPHA successfully captures multi-factor timing in simulations and among macro diversified hedge funds and flexible portfolio allocation mutual funds.

JEL classification: G11, G17, G23

Keywords: long-horizon timing, multi-factor timing, timing measures, hedge funds, mutual funds, performance evaluation, performance prediction

---

[a] Yuekun Liu, yuekun.liu@aalto.fi, School of Business, Aalto University, Espoo, 02150, Finland.
[b] Alexey Malakhov, amalakhov@walton.uark.edu, Sam M. Walton College of Business, University of Arkansas, Fayetteville, AR 72701, USA.
We are thankful to Jun Duanmu and William McCumber for the past collaboration on preliminary ideas. We thank Kelley Bergsma Lovelace, Ravi Jagannathan, Juha Joenväärä, Petri Jylhä, Matti Keloharju, Wayne Lee, Jonathan Lewellen, Xi Li, Matthijs Lof, Andrew Lynch, Imtiaz Mazumder, Peter Nyberg, Timothy Riley, Natasa Todorovic, Qing Yan, participants in the 2023 Financial Management Association European Conference, the 2023 Financial Management Association Annual Meeting and seminar participants at Aalto University, Ohio University, and the University of Arkansas for valuable discussions, comments, and suggestions.

1 Introduction

We focus on capturing multi-factor timing activity where portfolio managers anticipate macroeconomic trends and invest in the most profitable factor among multiple available factors. Realistically, such timing activity can only be implemented over relatively long horizons: trends, like a broad downturn or an upturn in a particular factor, are not matters of just a few months but potentially years. In fact, flexible asset allocation and macro investment styles in mutual and hedge funds rely on anticipating such trends. Haddad, Kozak, and Santosh (2020) confirm that successful multi-factor timing strategies may be implementable by investors over long horizons. On the other hand, Chernov, Lochstoer, and Lundeby (2022) demonstrate that linear factor models are not very good at pricing long-horizon returns, attributing that to the conditional factor dynamics. Whether active portfolio managers can successfully exploit the long-term time-varying nature of investable risk factors is an open question, motivating our focus on long-horizon multi-factor timing. Furthermore, while the potential and the possibility of successful long-horizon multi-factor timing have been documented in the literature, there are currently no measures with documented efficacy in capturing such activity by active portfolio managers.[1] In this paper, we first explore the efficacy of multiple timing measures in capturing long-horizon multi-factor timing with simulated timing data. We then apply the measures that proved to be effective in simulations to hedge and mutual fund data.

While multi-factor timing could be considered a straightforward concept, capturing successful multi-factor timing activity over long horizons is not trivial. For example, a factor with the most profitable long-term trend may not generate the highest monthly return among all factors every month in the trend. Moreover, the economic benefit of long-horizon multi-factor timing

[1] Given the focus of the extant literature on the single-period single-factor timing activity.

depends on the profitability of the optimal factor relative to other available factors over timing horizons. This, in turn, depends on the relative strength of the positive return trend in the optimal factor compared to other factors, which could vary across time. Perhaps, the most challenging aspect in capturing such timing activity is that portfolio managers make errors not only in the choice of an optimal factor, but also in the choice of the optimal timing horizon to follow that factor. Importantly, timing horizons chosen by portfolio managers are likely to be dynamic, depending on contemporaneous economic conditions. These timing horizons, being unobservable, are not likely to match any specific measurement horizon, making it impossible to deploy any reasonably specified structural model.[2] Therefore, we rely on multi-factor simulations and out-of-sample analysis to evaluate the efficacy of the multi-factor timing measures proposed in this study.

Single-factor timing has been first studied in pioneering works by Treynor and Mazuy (1966), Jensen (1972), and Henriksson and Merton (1981) in the context of market timing, proxied by a broad market index (for example, S&P 500). However, the existing literature mainly explores one-dimensional timing, while very few papers focus on active portfolio managers timing multiple factors.[3] On the other hand, the recent popularity of indexing and factor investing highlights the need for a comprehensive investigation of multi-factor timing.[4] Furthermore, it is prevalent in the timing literature to rely on the single-period timing framework, not paying due attention to timing activity over long horizons.[5]

---

[2] The mismatch between the investment horizon and the measurement horizon was considered in Jensen (1969, 1972), Fama (1972), Levhari and Levi (1977), Handa, Kothari, and Wasley (1989), Lee, Wu, and Wei (1990), Barberis (2000), Kamara, Korajczyk, Lou, and Sadka (2016), among others.

[3] For example, Aragon (2002), Comer (2006), and Duanmu, Malakhov, and McCumber (2018) attempt quantifying timing across multiple factors.

[4] See, for example, Ang (2014), Glushkov (2016), Arnott, Harvey, Kalesnik, and Linnainmaa (2019), Gupta and Kelly (2019), Haddad, Kozak, and Santosh (2020), Mateus, Mateus, and Soggiu (2020), Ehsani and Linnainmaa (2022), Ben-David, Franzoni, Kim, and Moussawi (2023).

[5] In traditional single-period timing literature, managers make factor investment decisions every single time period, so the only possible timing error could be a manager choosing a suboptimal factor. In long-term timing, in addition to the factor choice, managers also make decisions about specific times to switch into their chosen factor and how long

We propose a new indirect measure of long-horizon timing activity, D_ALPHA, building on the pioneering work of Jensen (1972). Jensen (1972) demonstrates that, in general, it is impossible to separate the market timing component of managerial performance in the context of a factor model, and Jensen (1968) alpha (alpha, hereafter) reflects both timing and security selection components of managerial performance. The concept of market timing involves active portfolio managers who can successfully anticipate market (or, in general, any factor) trends, and adjust their market portfolio exposure, beta, according to their forecast. Such time-varying beta exposures, if done successfully (i.e., high beta exposures during periods of good market performance and low beta exposures during periods of poor market performance), result in convexity in portfolio returns relative to the market factor. Jensen (1972) demonstrates that a linear factor model estimate of alpha for a successful timer would be upwardly biased from the value of alpha that only reflects security selection skills. Jensen (1972) interprets this upward bias in the linear factor model alpha as a reflection of timing skills.

Moreover, Jensen (1972) articulates that capturing timing activity is particularly difficult if the length of a timing horizon does not match the length of a measurement horizon. Figure 1 shows examples of mismatched timing and measurement horizons. If the measurement horizon is shorter than the timing horizon, there would be no timing changes in beta over the measurement horizon, and alpha would not reflect the timing component at all.[6] On the other hand, if the measurement horizon is longer than the timing horizon, there would be changes in beta over the

to follow the chosen factor. Notice that the decision on the specific length of the timing window is embedded in the decision about the specific time to switch into the next factor.

[6] In a similar spirit, though in a much different context, Lewellen and Nagel (2006) assume that CAPM parameter estimates are stable over short time windows and interpret them as conditional estimates in their tests of the conditional CAPM.

measurement horizon, and, according to Jensen (1972), alpha would reflect both timing and security selection components of managerial performance together.

Hence, we propose a new measure of long-horizon timing activity, D_ALPHA, as the difference between the long- and short-term alphas (fully described in Section 2). We conjecture for D_ALPHA to capture timing activity over horizons that approximately match the short-term time window. We interpret D_APLHA as an indirect measure of timing activity, given that timing activity would be captured not by a direct timing factor, but as the indirect result of the upward bias in long-horizon alphas.

We highlight the difficulty of capturing long-horizon multi-factor timing activity, which depends on the specific set of investable factors available for timing, through simulations. We simulate long-horizon multi-factor timing activity and compare in- and out-of-sample efficacy in capturing multi-factor timing activity of D_ALPHA along with multi-factor extensions of the Henriksson and Merton (1981) approach (HM approach, hereafter) and multi-factor conditional timing measures, introduced in Cederburg, O'Doherty, Savin, and Tiwari (2018) (COST measures, hereafter). In simulations, most considered measures successfully capture long-horizon multi-factor timing in-sample. Out-of-sample, D_ALPHA and most conditional COST measures successfully capture long-horizon multi-factor timing, but most of the multi-factor extensions of the HM measure fail to capture long-horizon multi-factor timing.

We then apply the proposed timing measures to hedge and mutual fund data, extending the literature on timing as well as performance attribution in mutual and hedge funds.[7] We find

[7] For mutual funds, for example, Jensen (1968, 1969, 1972), Fama (1972), Sharpe (1975, 1992), Merton (1981), Chang and Lewellen (1984), Henriksson (1984), Admati, Bhattacharya, Pfleiderer, and Ross (1986), Jagannathan and Korajczyk (1986), Lehmann and Modest (1987), Grinblatt and Titman (1994), Malkiel (1995), Ferson and Schadt (1996), Graham and Harvey (1996), Daniel, Grinblatt, Titman, and Wermers (1997), Busse (1999), Edelen (1999), Goetzmann, Ingersoll, and Ivkovic (2000), Moskowitz (2000), Wermers (2000), Bollen and Busse (2001, 2004), Chance and Hemler (2001), Chan, Chen, and Lakonishok (2002), Jiang (2003), Avramov and Wermers (2006), Comer (2006), Kosowski, Timmermann, Wermers, and White (2006), Goetzmann, Ingersoll, Spiegel, and Welch (2007),

evidence consistent with successful long-horizon multi-factor timing among macro diversified hedge funds and flexible portfolio allocation mutual funds. We conclude that, out of all considered measures, only D_ALPHA consistently captures long-horizon multi-factor timing out-of-sample in simulations and among hedge and mutual funds with investment styles that rely on long-term multi-factor timing.

2 Motivating Example and D_ALPHA Measure

2.1 Motivating Example

Consider timing the broad market index, S&P 500, over 24 months from March 2008 until February 2010, assuming the alternative of earning the risk-free rate of return by investing in the U.S. Treasury Bills. We assume that it is unrealistic to expect a hypothetical market timer to successfully anticipate market movements every month, but a successful long-horizon timer could anticipate the broad market downturn from March 2008 to February 2009, followed by the upturn from March 2009 to February 2010. Furthermore, we consider an imperfect market timer, who switches out of the Treasury Bills into the S&P 500 index not at the optimal switch time in March 2009, but two months later, in May 2009, as presented in Figure 2. Figure 2 displays cumulative excess returns for the market index and the imperfect market timer. While the imperfect timer did miss the optimal timing into the S&P 500 index by two months, it is clear from Figure 2 that this

---

Jiang, Yao, and Yu (2007), Kacperczyk, Sialm, and Zheng (2008), Mamaysky, Spiegel, and Zhang (2008), Bergstresser, Chalmers, and Tufano (2009), Cremers and Petajisto (2009), Chen, Ferson, and Peters (2010), Fama and French (2010), Ben-Rephael, Kandel, and Wohl (2012), Elton, Gruber, and Blake (2012), Amihud and Goyenko (2013), Fulkerson (2013), Hunter, Kandel, Kandel, and Wermers (2014), Kacperczyk, van Nieuwerburgh, and Veldkamp (2014), Bodnaruk, Chokaev, and Simonov (2018), Chuprinin and Sosyura (2018), among others.

For hedge funds, for example, Fung and Hsieh (1997, 2001, 2004), Agarwal and Naik (2000, 2004), Asness, Krail and Liew (2001), Aragon (2002), Fung, Xu and Yau (2002), Chen (2007), Chen and Liang (2007), Fung, Hsieh, Naik, and Ramadorai (2008), Bollen and Whaley (2009), Avramov, Kosowski, Naik, and Teo (2011), Bali, Brown, and Caglayan (2011, 2012, 2014), Titman and Tiu (2011), Cai and Liang (2012a, 2012b), Cao, Chen, Liang, and Lo (2013), Patton and Ramadorai (2013), Namvar, Phillips, Pukthuanthong, and Rau (2016), Agarwal, Green, and Ren (2018), Duanmu, Malakhov, and McCumber (2018), among others.

imperfect realistic timing activity results in obvious economic benefits and qualifies as successful long-horizon market timing.

As a motivating example, consider alphas from the single-factor model,

$$r_t - r_t^f = \alpha + \beta\left(r_t^{market} - r_t^f\right) + \varepsilon_t, \qquad (1)$$

for the imperfect timer from the overall 24 months and from three overlapping 12-month windows, rolling annually within the 24-month window, presented in Figure 3 and Table 1. Consistent with Jensen (1972), the 24-month positive alpha of 0.012 is the result of the single factor model, applied over the long 24-month horizon to the timing activity over shorter horizons.[8] Notice that the average of three 12-month alphas is only 0.007, reflecting a smaller average upward bias in the single factor alpha for the timing activity over 14- and 10-month horizons, measured over three overlapping 12-month horizons. It is important to emphasize the overlapping windows approach, given that timing horizons of portfolio managers are not observable, and their timing activity may not be well reflected by non-overlapping windows.

Following Jensen (1972), we conclude that the 24-month alpha exceeds the average of 12-month alphas due to the greater upward bias in alpha, when the measurement horizon exceeds the average timing horizon. Hence, we propose a new measure, D_APLHA = long-term alpha – average of short-term alphas: in this example, D_ALPHA = 0.012 – (0.000 + 0.016 + 0.006)/3 = 0.005.

2.2 D_APLHA Measure in a General Setting

In a general setting, D_ALPHA reflects the difference between the long-term alpha and the annual rolling average of overlapping short-term alphas within the long-term window, where the

[8] The timing activity in this example is the change in the market factor exposure, β, from 0 in March 2008 - April 2009, to 1 in May 2009 - February 2010. This corresponds to timing horizons of 14 and 10 months within the 24-month measurement horizon.

exact number of short-term time windows, N, depends on the difference in length between the long-term and the short-term windows. Specifically, D_ALPHA for portfolio manager i,

$$D\_ALPHA_i = \alpha_{i,long-term} - \frac{\alpha_{i,short-term\_1} + \cdots + \alpha_{i,short-term\_N}}{N}, \quad (2)$$

where $\alpha_{i,long-term}$ results from a general linear performance evaluation model over a long-term window, and $\alpha_{i,short-term\_1}, \ldots, \alpha_{i,short-term\_N}$ result from the same model over N short-term windows, rolling annually within the long-term window. The short-term window approximates the average timing horizon of portfolio managers, and we apply the overlapping short-term windows approach due to unobservability of the timing horizons of portfolio managers.

Importantly, our definition of D_ALPHA is very general and applies to any potential single or multi-factor linear performance evaluation model over any statistically feasible time horizon.[9] Following Jensen (1972), we conjecture that timing activity over horizons that are close to the length of the short-term windows on average would result in greater positive alphas over the long-term window while generating lower average alphas over the short-term windows.[10] We interpret D_APLHA as an indirect measure of timing activity over any potential set of factors, given that successful timing would be captured not by a direct timing factor, but as an indirect result of the upward bias in long-term alphas due to timing activity. It is important to emphasize that D_ALPHA reflects factor timing from the perspective of generating alpha long-term. In applications to hedge and mutual funds, it may reflect other dimensions of active portfolio management that may produce time-varying short-term alphas while generating positive long-term alphas.

[9] The determination of the appropriate length of long-term and short-term time windows depends on the research objectives and the availability of data. The more factors are included in the baseline factor model, the more data points are required for a meaningful statistical analysis, necessitating longer time windows.

[10] Under the assumption that alpha generation through security selection is stationary over time.

Next, we introduce and discuss the full set of timing measures that we consider as potential candidates for capturing long-horizon multi-factor timing.

3 Multi-Factor Timing: Measures and Factors

We consider all the timing measures in this paper based on a general multi-factor performance evaluation model for a fund i,

$$r_{i,t} - r_t^f = \alpha_i + \sum_{j=1}^{N} \beta_i^j Factor_t^j + \varepsilon_{i,t}\,. \qquad (3)$$

Given our focus on long-horizon timing, we calculate all measures over the 48-month time window. Following the logic in Jensen (1972), highlighted in Figure 1, the 48-month long-term window allows us to capture timing activity over horizons around the short-term window of 24 months and also accommodates a sufficiently large number of factors in multi-factor timing regressions. While our choice of the 24-month short-term window reflects the shortest time window over which a meaningful statistical analysis based on monthly data may be executed, our analysis may be easily extended to different length time windows as a robustness check, based on the availability of data. For example, with monthly data, a 30-month window may be considered; with weekly data, considering 12- and 6-month windows could be reasonable.

3.1 Unconditional Measures of Long-Horizon Multi-Factor Timing Activity

We apply the D_APLHA formula (2) from the previous section to calculate D_ALPHA in the context of model (3) with short-term windows of 24 months and the long-term window of 48 months as follows:

$$D\_ALPHA_i = \alpha_{i,48} - \frac{\alpha_{i,24-1} + \alpha_{i,24-2} + \alpha_{i,24-3}}{3}\,, \qquad (4)$$

where $\alpha_{i,48}$ results from model (3) over the 48-month window, and $\alpha_{i,24-1}$, $\alpha_{i,24-2}$, $\alpha_{i,24-3}$ result from model (3) over three overlapping 24-month windows, rolling annually within the above-mentioned 48-month window.[11] As mentioned in Section 2, D_ALPHA is very general in its definition, and it can be applied to any linear single- or multi-factor evaluation model over any statistically feasible time horizon. We interpret D_ALPHA as an indirect measure of timing activity over average time horizons of 24 months.

Extending classic Treynor and Mazuy (1966) and Henriksson and Merton (1981) approaches in a multi-factor setting is not trivial, given potential cross-correlations among the factors.[12] We consider the HM approach of interpreting successful market timing as successfully exercising a call option on the market portfolio, max(0, market factor), to be suitable for direct multi-factor extensions, as it assumes a manager to be fully invested in the timed factor. Given that a manager cannot be fully invested in more than one factor at the same time, cross-correlations among the factors are not likely to present additional complications upon adding HM timing factors, max(0, $Factor_1$),…, max(0, $Factor_N$), to the regression (3). However, interpreting multiple HM timing coefficients is not trivial, given that most of the available factors are not likely to be timed into during any given 48-month horizon window. Imperfect timing, when managers make errors in the choice of the optimal factor and in the optimal time to switch into the optimal factor, further complicates the interpretation of multiple HM timing coefficients. Thus, we consider three

[11] It is possible to argue that $\alpha_{i,24-2}$ could be weighed lower than $\alpha_{i,24-1}$ and $\alpha_{i,24-3}$ in the D_ALPHA calculation (4), given that it is based on observations partially included into calculating of $\alpha_{i,24-1}$ and $\alpha_{i,24-3}$. On the other hand, $\alpha_{i,24-1}$ and $\alpha_{i,24-3}$ could be less reliable in reflecting long-horizon timing, given that they may misrepresent timing activity that either starts before or extends beyond the 48-month window. For this reason, we apply equally weighted average to $\alpha_{i,24-1}$, $\alpha_{i,24-2}$, and $\alpha_{i,24-3}$ in calculating D_ALPHA in (4). For robustness, we also calculate a weighted version of D_ALPHA with $\alpha_{i,24-1}$, $\alpha_{i,24-2}$, $\alpha_{i,24-3}$ weights as 0.375, 0.25, 0.375. The results are qualitatively and quantitatively similar to the equally weighted approach in (4) and are available upon request. Also, D_ALPHA may be calculated with non-overlapping short-term windows as an additional robustness check.

[12] The Treynor and Mazuy (1966) approach relies on including the square term for the timed factor into a linear factor model. Its multi-factor extension, introduced in Lehmann and Modest (1987), includes cross-products of the factors in addition to squared terms for all factors. Unfortunately, as articulated in Chen, Ferson, and Peters (2010), such an approach imposes severe data requirements, and it cannot be meaningfully executed in our setting with 48 monthly observations.

multi-factor HM extensions with different interpretations for measuring multi-factor timing activity.

First, we consider MAX_ALL measure as the regression coefficient γ_{MAX_ALL} for the call option on the optimal factor choice at any given time, $\max(0, Factor_t^1, Factor_t^2, \dots, Factor_t^N)$, modifying model (3) as follows:

$$r_{i,t} - r_t^f = \alpha_i + \sum_{j=1}^{N} \beta_i^j Factor_t^j + \gamma_{i,MAX\_ALL} \max(0, Factor_t^1, Factor_t^2, \dots, Factor_t^N) + \varepsilon_{i,t}. \quad (5)$$

We interpret MAX_ALL as a measure of success in timing the most profitable factor every month.

Second, we modify model (3) by adding call options on all individual factors as follows:

$$r_{i,t} - r_t^f = \alpha_i + \sum_{j=1}^{N} \beta_i^j Factor_t^j + \sum_{j=1}^{N} \gamma_i^j \max\left(0, Factor_t^j\right) + \varepsilon_{i,t}. \quad (6)$$

We consider MAX_SINGLE measure based on model (6) as

$$MAX\_SINGLE_i = \max(\gamma_i^1, \gamma_i^2, \dots, \gamma_i^N). \quad (7)$$

We interpret MAX_SINGLE as a measure of timing activity in the most successfully timed single factor.

Finally, we consider the last extension of the HM measure, SUM_ALL as

$$SUM\_ALL_i = \sum_{j=1}^{N} \gamma_i^j, \quad (8)$$

with γ_i^j coefficients from the model (6). We interpret SUM_ALL as a measure of overall timing activity across all factors.

3.2 Conditional Measures of Long-Horizon Multi-Factor Timing Activity

While we interpret D_ALPHA as an unconditional performance measure, it can also be interpreted from a conditional performance evaluation perspective.[13] Conceptually, any active portfolio management is a result of decisions by portfolio managers conditional on contemporaneous information. Unfortunately, specific information sets that the managers base their investment decisions on, and their decision rules are unobservable. Hence any conditional approach that incorporates publicly observable instrument variables into factor regressions could be prone to misspecification in case the instruments do not adequately capture contemporaneous information utilized by portfolio managers. This particularly applies to long-horizon factor timing, when the relevance of specific contemporaneous information heavily depends on the personal judgment of the managers. For this reason, we focus on conditional timing measures that are based on past decisions by portfolio managers, reflected by beta exposures, introduced in Cederburg, O'Doherty, Savin, and Tiwari (2018) (COST, hereafter).

Following COST, we rely on the past beta exposures of active portfolio managers as instrumental variables in the conditional timing evaluation. As we focus on long-horizon timing over the 24-month horizon, we use the past 24-month beta estimates as instruments in the conditional model for a fund *i* as follows.

$$r_{i,t} - r_t^f = \alpha_i^C + \sum_{j=1}^{N} (\lambda_{i,0}^j + \lambda_{i,1}^j \beta_{i,t-1}^j) Factor_t^j + \varepsilon_{i,t}, \qquad (9)$$

where $\beta_{i,t-1}$ is estimated via the unconditional model (3) over the 24 months preceding a given *t*, i.e. $\tau \in [t-1, \ldots, t-24]$ in

[13] For example, following Boguth, Carlson, Fisher, and Simutin (2011), D_ALPHA can be interpreted as a conditional measure with a positive "overconditioning" bias. However, the focus of Boguth, Carlson, Fisher, and Simutin (2011) is on correcting the "overconditioning" due to time-varying betas, while the focus of D_ALPHA is on utilizing time-varying beta exposures and the resulting difference between short-term and long-term alphas to reflect factor timing over horizons comparable to the length of the short-term window.

$$r_{i,\tau} - r_\tau^f = \alpha_i + \sum_{j=1}^{N} \beta_{i,t-1}^j Factor_\tau^j + \varepsilon_{i,\tau}. \qquad (10)$$

In the conditional model (9), α_i^C represents conditional alpha (COND_ALPHA, hereafter) and

$$\beta_{i,t}^{j,C} = \lambda_{i,0}^j + \lambda_{i,1}^j \beta_{i,t-1}^j \qquad (11)$$

represents conditional betas for a fund *i*.

COST defines the conditional factor timing measure for a fund *i* and a factor *j* as $FT_i^j = cov(\beta_{i,t}^{j,C}, Factor_t^j)$. Following that, we consider three conditional measures of multi-factor timing. The first one, COND_MAX_FT, is a multi-factor extension of FT_i^j, defined as

$$COND\_MAX\_FT_i = \max(FT_i^1, FT_i^2, \ldots, FT_i^N). \qquad (12)$$

We borrow the second and the third conditional measures of multi-factor timing, total factor timing (COND_TOTAL_FT) and total volatility timing (COND_VOL), directly from COST as follows:

$$COND\_TOTAL\_FT_i = \sum_{j=1}^{N} FT_i^j, \qquad (13)$$

$$COND\_VOL_i = \sum_{j=1}^{N} (\overline{\beta_{i,t}^{j,C}} - \beta_i^j)\, \overline{Factor_t^j}, \qquad (14)$$

where β_i^j represents unconditional betas for a fund *i* from model (3).

3.3 Factor Models for Hedge and Mutual Fund Performance Evaluation

While the introduced timing measures are very general, we apply them in the context of different factor models for hedge and mutual fund performance evaluation. Given the focus of our analysis on long-horizon multi-factor timing, we employ performance evaluation models with factors that can be interpreted as investable portfolios suitable for realistic long-term timing by active portfolio managers.

For hedge funds, we use the Fung and Hsieh (2004) eight-factor performance evaluation model over 300 months from January 1994 to December 2018 as follows:

$$r_{i,t} - r_t^f = \alpha_i + \beta_i^1 SP500_t + \beta_i^2 EM_t + \beta_i^3 10Y_t + \beta_i^4 SizeSpread_t + \beta_i^5 CreditSpread_t + \beta_i^6 BondTrend_t + \beta_i^7 ComTrend_t + \beta_i^8 FxTrend_t + \varepsilon_{i,t}\,, \quad (15)$$

where r_i is the monthly return of a potential timer i, r^f is the risk-free rate proxied by the monthly return of the 30-day U.S. Treasury bill. $SP500$ is the market risk premium proxied by the excess return of the S&P 500 index. EM is the emerging market risk factor proxied by the excess return of the MSCI Emerging Market index. $SizeSpread$ is an equity-oriented risk factor, quantified by the Russell 2000 index total return minus the S&P 500 index total return. *10Y* and $CreditSpread$ are bond-oriented risk factors. *10Y* is proxied by the excess return of the 10-year U.S. Treasury Bond portfolio from CRSP. $CreditSpread$ is calculated as the total return on the S&P U.S. Investment Grade Corporate Bond BBB index minus the total return on the Fama U.S. Treasury bond portfolio from CRSP. $BondTrend$, $ComTrend$, and $FxTrend$ are excess returns on trend following factors constructed of look-back straddles on futures contracts of bonds, commodities, and currencies, respectively.[14]

For mutual funds, we rely on the following five-factor performance evaluation model over 300 months from January 1994 to December 2018:

$$r_{i,t} - r_t^f = \alpha_i + \beta_i^1 Market_t + \beta_i^2 Size_t + \beta_i^3 Value_t + \beta_i^4 Mom_t + \beta_i^5 10Y_t + \varepsilon_{i,t}\,, \quad (16)$$

where r_i is the monthly return of a potential timer i, r^f is the risk-free rate proxied by the monthly return of the 30-day U.S. Treasury bill. $Market$ is the market risk premium proxied by the excess return of the S&P 500 index. $Size$ is the size equity-oriented risk factor, quantified by the Russell

[14] All return data are from Bloomberg, Morningstar, and CRSP, while trend-following factors are from David Hsieh's website (https://faculty.fuqua.duke.edu/~dah7/HFData.htm).

2000 index total return minus the S&P 500 index total return. $Value$ is the value minus growth equity risk factor, quantified by the difference in returns between Russell 3000 Value and Russell 3000 Growth indices.[15] Mom is the momentum risk factor, quantified by the momentum index from Kenneth French data library.[16] $10Y$ is the bond-oriented risk factor, proxied by the excess return of the 10-year U.S. Treasury Bond portfolio from the Center for Research in Security Prices (CRSP).

The mutual fund performance evaluation model (16) is based on the model from Cremers, Petajisto, and Zitzewitz (2013), utilizing investable proxies for commonly known risk factors of market, size, value, and momentum,[17] along with the investable proxy for the bond risk factor, $10Y$, which is included to accommodate common mutual fund timing strategies of flexible portfolio allocation between stocks and bonds, depending on market forecasts of portfolio managers.

As mentioned above, we interpret the factors in models (15) and (16) as investable time-varying risk factors suitable for long-term timing by active portfolio managers. The difference in the set of factors between models (15) and (16) reflects the difference in the space of investable time-varying risk factors, potentially available for multi-factor timing, between hedge and mutual fund managers.

We rely on investable factors from models (15) and (16) as the basis for simulations and tests that follow.

[15] All index data are from Morningstar.
[16] Available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
[17] These risk factors were considered in Fama and French (1993), Jegadeesh and Titman (1993), and incorporated into evaluating mutual fund performance in Carhart (1997).

4 Long-Horizon Multi-Factor Timing: Simulations

As mentioned in the introduction, capturing successful multi-factor timing activity over long horizons is not trivial. For example, a factor with the most profitable long-term trend may not generate the highest monthly return among all factors every month in the trend. Imperfect long-horizon timing, when portfolio managers make mistakes in the timing of factor switches and the choice of factors, further complicates capturing long-horizon multi-factor timing activity with statistical and economic significance. Finally, economic benefits of long-horizon multi-factor timing depend on the profitability of the optimal factor relative to other available factors over long timing horizons. Hence, the efficacy of the proposed long-horizon timing measures depends on the specific set of investable risk factors potentially available for timing. Therefore, we first evaluate the efficacy of the proposed measures via the simulation process based on sets of factors for hedge and mutual funds from models (15) and (16). The objective of this simulation is to establish benchmarks and compare the efficacy of proposed multi-factor timing measures in the context of timing different sets of factors.

4.1 Simulated Long-Horizon Multi-Factor Timing with Hedge Fund Factors

We consider factor data from the Fung and Hsieh (2004) eight-factor model (15) from January 1994 to December 2018 and construct the optimal multi-factor timing pattern with an average timing horizon of 23.08 months.[18] Then we apply the proposed timing measures to 90,000 simulated portfolio managers (or imperfect timers), who attempt to follow the optimal timing pattern with various levels of accuracy in their multi-factor timing decisions. Thus, we benchmark our measures in the context of ex-ante known timing abilities applied to eight-factor data from

[18] The average timing horizon of 23.08 months in the optimal eight-factor timing pattern is close to the 24-month timing horizon for calculating D_ALPHA in (4). D_ALPHA reflects the difference between the 48-month alpha and the average of 24-month alphas.

model (15). Most important, this simulation allows us to quantify and compare the out-of-sample efficacy of our proposed measures.

We construct the optimal long-term timing pattern by considering factor timing switch points based on the future return data. Specifically, at any potential switch point, we calculate end values from investing $1 in each factor over the next 12-24 months, switching into the highest end value factor. After the switch, we hold the new factor over the time period that generated the highest end value and then consider switching factors again. In case the highest end value over the next 12-24 months is generated by the currently held factor, we hold the current factor for another month, and consider switching factors next month. Afterward, we manually review the optimal time pattern to ensure that it produces the highest possible end value among other possible timing patterns with the same average timing horizon. In the end, following the optimal timing pattern yields $6,605.72 in December 2018 based on $1 investment in January 1994, compared to $8.77, which is the highest value based on $1 invested in any single factor.[19] The optimal timing pattern is presented in Figure 4.

Next, we simulate 90,000 portfolio managers (or imperfect timers) who follow the optimal timing pattern with various levels of accuracy in their long-term multi-factor timing decisions. In our simulation, imperfect timers can make errors in both timing of switching into the optimal factor and in the optimal factor selection. We simulate three levels of accuracy in timing decisions and three levels of accuracy in factor selection: combined, this yields nine possible combinations in timing accuracy combined with factor selection accuracy; we simulate 10,000 timers for each of these combinations. Specifically, we simulate timing errors by allowing the factor switch time within N months prior to the optimal switch time. The exact month within the N-month error is

[19] The highest end value of $8.77 in December 2018 among all factors was generated by investing $1 in January 1994 in the S&P 500 index factor.

simulated via uniform distribution from zero to N. The three simulated levels in timing accuracy correspond to timing errors within N = 2, 4, and 6 months. Factor selection accuracy is simulated by assigning one of the three accuracy levels to the probability of selecting the optimal factor at timing switches, with the complementary probability assigned to the remaining non-optimal factors.[20] We simulate the optimal factor selection accuracy at each switch point with probabilities of 90%, 60%, and 30%.

We focus our analysis on seven proposed measures of long-horizon multi-factor timing activity: the indirect measure, D_ALPHA, plus three unconditional multi-factor extensions of the HM measure, MAX_ALL, MAX_SINGLE, SUM_ALL, and three COST conditional measures, COND_MAX_FT, COND_TOTAL_FT, COND_VOL. We also consider the classic performance measure, Jensen (1968) alpha (ALPHA, hereafter), and COST-based conditional alpha, COND_ALPHA. The inclusion of ALPHA and COND_ALPHA in our analysis is essential for three reasons. First, as articulated in Jensen (1972), long-term ALPHA reflects both timing and security selection components of managerial performance, so it could be interpreted as a timing measure in the context of our timing simulation.[21] Second, our proposed indirect long-term timing measure, D_ALPHA, is the derivative measure based on ALPHA. Third, as mentioned earlier, D_ALPHA could be interpreted from a conditional perspective. Hence it is critical to demonstrate that D_ALPHA captures a distinct aspect of long-term timing activity, not reflected in ALPHA and COND_ALPHA.

Table 2 presents in-sample summary statistics for each measure.[22] For each simulated timer, we calculate each measure over four-year windows throughout the sample period of 25 years,

[20] The exact non-optimal factor is selected at random with equal probabilities from all non-optimal factors.
[21] However, it is important to stress that interpreting ALPHA specifically as a timing measure is limited to the context of our simulations, given that timing is the only simulated dimension of active portfolio management.
[22] Standard deviations are presented in the Appendix Table A1.

which yields 21 observations for each unconditional measure, D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, and 19 observations for each conditional measure, COND_MAX_FT, COND_TOTAL_FT, COND_VOL, COND_ALPHA, given that calculating the conditional measures requires 24-month history data to calculate past 24-month beta exposures. For unconditional measures, this results in 21,000 (21*10,000 timers) observations for each of the nine accuracy options for imperfect timers and in 19,000 (19*10,000 timers) observations for conditional measures.

All measures, except SUM_ALL, COND_MAX_FT, COND_VOL, display a uniform pattern reflecting the accuracy of long-term five-factor timing activity: among imperfect timers, when one dimension of the timing skill is fixed, measure values monotonically decrease as the other dimension of the timing skills goes down.[23] Take D_ALPHA as an example. Its highest value is achieved for the optimal timing pattern. Among the imperfect timers, when the timing accuracy is fixed, D_ALPHA values monotonically decline as the factor selection accuracy decreases. When the factor selection accuracy is fixed, the values of D_ALPHA monotonically decrease as the timing accuracy goes down.

Next, we proceed with out-of-sample tests to quantify potential economic benefits to investors employing our measures. We form portfolios based on top and bottom deciles of in-sample values for each measure. For unconditional measures, D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, each portfolio is initially formed on December 31, 1997 based on ranked measure values over preceding 48 months. For conditional measures, COND_MAX_FT, COND_TOTAL_FT, COND_VOL, COND_ALPHA, each portfolio is initially formed on December 31, 1999 based on ranked measure values over preceding 48 months,

[23] COND_VOL displays a strong pattern with respect to factor selection accuracy.

conditional on preceding 24-month betas. We invest \$1 in each timer within a portfolio at the beginning of the year, rebalance the portfolio annually based on the updated ranking with respect to in-sample four-year window measure values, and track its performance until December 31, 2018. Thus, each portfolio for the unconditional measures has a time series of 252 monthly observations and 228 observations for the conditional measures. For each portfolio, we use the resulting time series to calculate the Fung and Hsieh (2004) eight-factor model (15) alpha with its t-value, the mean excess return, Sharpe ratio, and the End Value.[24] Table 3 shows the out-of-sample performance of top and bottom portfolios. The top portfolio comprises timers with the highest in-sample measure values in the top decile, and the bottom portfolio comprises the lowest in the bottom decile. The Top-Bottom portfolio is equivalent to taking a long position in the top portfolio and a short position in the bottom portfolio. Among all proposed measures, D_ALPHA, ALPHA, COND_MAX_FT, COND_VOL, and COND_ALPHA provide economic and statistical separation in the top and bottom portfolios, confirming that these measures provide valuable benefits to investors.[25] However, due to volatility in monthly factor performance and simulated timing and factor selection errors, MAX_SINGLE, SUM_ALL, and COND_TOTAL_FT measures do not deliver statistically significant separation out-of-sample. We conclude that D_ALPHA, ALPHA, COND_MAX_FT, COND_VOL, and COND_ALPHA can separate timers with various accuracy skills and provide economic benefits to investors in the context of long-horizon timing eight factors from model (15).

[24] The end value is the dollar value as of December 2018 based on the initial \$1 investment on December 31, 1997 for the unconditional measures, and on the initial \$1 investment on December 31, 1999 for the conditional measures.
[25] Note that both top and bottom portfolios may produce economically and statistically significant alphas, because all simulated timers possess varied degrees of timing skills. The significant performance separation in the top-bottom portfolio captures the difference in the skill level.

Last, we examine whether the measures that demonstrate predictive powers in out-of-sample tests in Table 3, D_ALPHA, ALPHA, COND_MAX_FT, COND_VOL, and COND_ALPHA, capture distinct aspects of long-term multi-factor timing activities. Table 4 provides the mutually exclusive portfolio analysis, where mutually exclusive portfolios isolate the top timers unique to each measure. Specifically, we first form top decile portfolios sorted by one measure and then exclude timers that are also in the top decile of the other measure. For example, "D_ALPHA Excl. ALPHA" includes timers present in the top decile of the D_ALPHA-sorted portfolio but not present in the top decile of the ALPHA-sorted portfolio. Table 4 shows the out-of-sample performance of these isolating portfolios. We observe that the performance metrics for the isolating portfolios are close to that of the original top portfolios in Table 3. In other words, the performance of the top portfolio from each measure is not impacted by the timers identified by other measures. Thus, we confirm that D_ALPHA, ALPHA, COND_MAX_FT, COND_VOL, and COND_ALPHA measures capture distinct aspects of long-horizon timing eight factors from the Fung and Hsieh (2004) model (15).[26]

We conclude that, out of the considered multi-factor timing measures, D_ALPHA, ALPHA, COND_MAX_FT, COND_VOL, and COND_ALPHA demonstrate out-of-sample efficacy and distinctiveness in reflecting different aspects of long-horizon timing eight factors from the Fung and Hsieh (2004) model (15).

4.2 Simulated Long-Horizon Multi-Factor Timing with Mutual Fund Factors

Now we consider factor data from the five-factor model (16) for long-horizon multi-factor simulation, following the methodology of the eight-factor simulation in the section above. As

[26] It is essential to confirm that D_ALPHA captures a unique aspect of multi-factor timing activity, distinct from those captured by ALPHA and COND_ALPHA, given that D_ALPHA is the alpha-based measure.

mentioned above, the objective of this simulation is to compare the efficacy of proposed long-term multi-factor timing measures in the context of timing different sets of factors.

First, we construct the optimal five-factor timing pattern from January 1994 to December 2018 with an average timing horizon of 21.43 months.[27] The optimal timing pattern yields $372.90 in December 2018 based on $1 investment in January 1994, compared to $8.77, which is the highest value based on $1 invested in any single factor.[28] The optimal timing pattern is presented in Figure 5.

We then consider 90,000 simulated timers, described in the previous section, timing the five factors from model (16). The in-sample simulation results for D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, COND_MAX_FT, COND_TOTAL_FT, COND_VOL, and COND_ALPHA are presented in Table 5. D_ALPHA, MAX_ALL, MAX_SINGLE, ALPHA, and COND_ALPHA measures display a uniform pattern reflecting the accuracy of long-term five-factor timing activity: among imperfect timers, when one dimension of the timing skill is fixed, measure values monotonically decrease as the other dimension of the timing skill goes down. COND_MAX_FT and COND_TOTAL_FT display a strong pattern with respect to factor selection accuracy.

Next, we proceed with out-of-sample tests to quantify potential economic benefits to investors employing our measures. Table 6 provides the out-of-sample performance analysis for each measure, following the methodology from the previous section. The results indicate that D_ALPHA, MAX_ALL, ALPHA, COND_MAX_FT, COND_TOTAL_FT, and COND_ALPHA

[27] The average timing horizon of 21.43 months in the optimal five-factor timing pattern is close to the 24-month timing horizon for calculating D_ALPHA in (4). D_ALPHA reflects the difference between the 48-month alpha and the average of 24-month alphas.

[28] The highest end value of $8.77 in December 2018 among all factors was generated by investing $1 in January 1994 in the S&P 500 index factor.

measures generate economic and statistical separation between the top and bottom portfolios. However, due to volatility in monthly factor performance and simulated timing and factor selection errors, MAX_SINGLE, SUM_ALL, and COND_VOL measures do not deliver statistically significant separation out-of-sample. We conclude that D_ALPHA, MAX_ALL, ALPHA, COND_MAX_FT, COND_TOTAL_FT, and COND_ALPHA can separate timers with various accuracy skills and provide economic benefits to investors in the context of long-horizon timing five factors from model (16).

Last, we examine whether the measures that demonstrate predictive power in out-of-sample tests in Table 6, D_ALPHA, MAX_ALL, ALPHA, COND_MAX_FT, COND_TOTAL_FT, and COND_ALPHA, capture distinct aspects of multi-factor timing activities. Table 7 provides the out-of-sample performance analysis for mutually exclusive portfolios, following the methodology from the previous section. The results indicate that performance metrics for the isolating portfolios are comparable to that of the top portfolios shown in Table 6. This confirms that D_ALPHA, MAX_ALL, ALPHA, COND_MAX_FT, COND_TOTAL_FT, and COND_ALPHA measures capture distinct aspects of long-horizon timing five factors from model (16).

We conclude that, out of the considered multi-factor timing measures, D_ALPHA, MAX_ALL, ALPHA, COND_MAX_FT, COND_TOTAL_FT, and COND_ALPHA demonstrate out-of-sample efficacy and distinctiveness in reflecting different aspects of long-horizon timing five factors from model (16).

4.3 Timing Simulations Comparison

Overall, there is high consistency in eight- and five- factor long-horizon timing simulation results, with D_ALPHA, ALPHA, COND_MAX_FT, and COND_ALPHA demonstrating high out-of-sample efficacy and distinctiveness in both simulations. It is notable that out-of-sample results for D_ALPHA, ALPHA, and COND_ALPHA measures for top decile portfolios appear to be quantitatively larger in the eight-factor simulation compared to the five-factor simulation. We conjecture that this could be due to the larger set of available factors in the eight-factor model (15), as well as considerably higher volatilities among derivative trend-following factors, $BondTrend$, $ComTrend$, and $FxTrend$, included in the eight-factor model (15), compared to volatilities among the factors in the five-factor model (16).[29]

We proceed with the analysis of long-term multi-factor timing activity of hedge and mutual fund managers by applying timing measures that exhibit out-of-sample predictive power in simulations above to hedge and mutual fund data. However, it is important to emphasize that the efficacy of ALPHA and COND_ALPHA as specifically timing measures has been established in the context of timing-only simulations based on specific factors included in models (15) and (16), with timing being the only simulated dimension of active portfolio management. It is possible for real-life portfolio managers to time other investable factors not included in models (15) and (16). Furthermore, in real-life applications to hedge and mutual funds, ALPHA and COND_ALPHA would also reflect other aspects of active portfolio management unrelated to timing, such as security selection, arbitrage, derivative speculation, and risk management, among others. This will complicate interpretations of APLHA and COND_ALPHA as timing measures in real-life applications.

[29] This is consistent with Jensen (1972) conclusion of positive relationship between the magnitude of the upward bias in alpha due to timing activity and the volatility of the timed factor.

5 Multi-Factor Timing Activity among Hedge Fund Managers

In the eight-factor long-horizon timing simulation in Section 4.1, D_ALPHA, ALPHA, COND_MAX_FT, COND_VOL, and COND_ALPHA exhibit predictive power in performance separation out-of-sample. In this section, we apply these measures, based on the eight-factor Fung and Hsieh (2004) model (15), to hedge fund data to investigate long-horizon multi-factor timing skills of hedge fund managers. We focus on macro diversified hedge funds, as their "investment process is predicated on movements in underlying economic variables and the impact these have on equity, fixed income, hard currency and commodity markets, (…) designed to identify opportunities in markets exhibiting trending or momentum characteristics" is most consistent with long-horizon multi-factor timing.[30] We examine whether managers with superior long-horizon multi-factor timing skills, according to our measures, deliver significantly superior performance in- and out-of-sample. Ultimately, we explore whether employing our measures can provide economic value to real-world investors in hedge funds.

5.1 Hedge Fund Data

We obtain hedge fund data from Bloomberg for the period from January 1994 to December 2018. The data includes both live and defunct hedge funds that were acquired, liquidated, or ceased to report during the sample period, mitigating the survivorship bias. Our analysis focuses on macro diversified hedge funds, given that these hedge fund managers are most likely to engage in long-term multi-factor timing activities. We eliminate the first 24 monthly observations for all funds to partially offset the effects of backfill bias, following the methodology in Jagannathan, Malakhov, and Novikov (2010, 2013) and Titman and Tiu (2011). Since reliable hedge fund data without survivorship bias only starts in 1994, eliminating the first 24 monthly observations for all funds

[30] Quoted from HFR hedge fund style definitions at https://www.hfr.com/hfr-hedge-fund-strategy-definitions-macro.

implies that we only use hedge fund data starting from January 1996 in our subsequent analysis. Our sample contains 389 macro diversified hedge funds with 17,672 observations over the time period from 1996 to 2018. Table 8 reports summary statistics of hedge fund returns, fees, investor liquidity measures, and fund longevity.

5.2 Hedge Fund In-Sample Analysis

In our hedge fund analysis, we focus on the measures that demonstrate predictive power in the long-horizon eight-factor timing simulation, D_ALPHA, ALPHA, COND_MAX_FT, COND_VOL, and COND_ALPHA. Table 9 presents in-sample measure values, calculated from net hedge fund returns, following the rolling window methodology from Section 4. To interpret in-sample variation in long-horizon multi-factor timing skills across macro diversified hedge fund managers, we provide the mean of each measure within each quartile of in-sample values. Q4 represents the average measure value in the top quartile, and Q1 represents the average value in the bottom quartile.

The top quartile values of all measures are significantly higher than those in the bottom quartile, suggesting significantly better multi-factor timing skills among the top macro diversified hedge fund managers compared to the bottom ones.

5.3 Hedge Fund Out-of-Sample Analysis

We now consider whether employing our measures can provide economic value to investors. In other words, we explore whether our measures can identify skilled managers with statistically and economically superior performance out-of-sample. In Table 10, we follow the methodology for out-of-sample tests from Section 4 based on top and bottom quartile portfolios. Unlike simulated timers in Section 4, hedge funds have various life spans. In our analysis, when a fund stops reporting, we assume the fund is defunct and redistribute the remaining capital in the fund equally among surviving portfolio funds. We report attrition rates as average annual rates at which hedge funds stop reporting.

For macro diversified hedge funds, the top quartile D_ALPHA portfolio generates a statistically significant alpha of 0.272, and it outperforms the bottom quartile D_ALPHA portfolio by a statistically significant out-of-sample alpha of 0.417.[31] On the contrary, ALPHA, COND_MAX_FT, COND_VOL, and COND_ALPHA generate qualitatively opposite out-of-sample results, with the bottom portfolios performing better than the top portfolios.

We conclude that the out-of-sample results for D_ALPHA are consistent with the presence of persistent skill in long-horizon multi-factor timing activity among macro diversified hedge fund managers.[32] We further discuss the results in Section 7.

6 Multi-Factor Timing Activity among Mutual Fund Managers

In the five-factor long-horizon timing simulation in Section 4.2, D_ALPHA, MAX_ALL, ALPHA, COND_MAX_FT, COND_TOTAL_FT, and COND_ALPHA exhibit predictive power in performance separation out-of-sample. In this section, we apply these measures, based on the five-factor model (10), to mutual fund data to investigate long-horizon multi-factor timing skills of mutual fund managers. We focus on flexible portfolio allocation mutual funds, as their investment objectives to "allocate investments across various asset classes, including domestic common stocks, bonds, and money market instruments with a focus on total return" are most consistent with long-horizon multi-factor timing.[33] We examine whether managers with superior long-horizon

[31] Out-of-sample alphas of 0.272 and 0.417 are based on monthly returns, equivalent to annualized alphas of 3.264% and 5.004%.
[32] Considering the full sample of hedge funds, D_ALPHA does not exhibit out-of-sample predictive power, which is not surprising, given that hedge funds follow a wide variety of investment styles, most of which do not involve timing. These results are available upon request.
[33] Quoting from CRSP Lipper Objective and Classification Codes at https://www.crsp.org/products/documentation/lipper-objective-and-classification-codes.

multi-factor timing skills, according to our measures, deliver significantly superior performance in- and out-of-sample.

6.1 Mutual Fund Data

We use the CRSP Survivor-Bias-Free Mutual Fund Database to construct our mutual fund sample. The sample period spans the period from January 1994 to December 2018. Our analysis focuses on mutual funds with CRSP Lipper objective classifications of Flexible Portfolio Funds, Global Flexible Portfolio Funds, and Alternative Global Macro Funds, given that these mutual fund investment styles are most consistent with long-term multi-factor timing activities.[34] For simplicity, we refer to this subsample as "flexible portfolio allocation mutual funds". To address incubation bias (Evans, 2010), we remove funds from the sample until they are at least two years old and until they reach at least $15 million in assets. We collapse all fund share classes into a single fund using the WFICN variable provided by MFLINKS. Fund characteristics and returns are asset-weighted averages of share classes, and the assets are aggregated across all share classes. Our sample contains 343 unique funds with 46,925 observations. Table 11 reports summary statistics for the sample of flexible portfolio allocation mutual funds of total net assets (TNA), expense ratios, turnover rates, and returns.

6.2 Mutual Fund In-Sample Analysis

In our mutual fund analysis, we focus on the measures that demonstrate out-of-sample predictive power in the five-factor long-horizon timing simulation, D_ALPHA, MAX_ALL, ALPHA, COND_MAX_FT, COND_TOTAL_FT, and COND_ALPHA. Table 12 presents in-

[34] For example, according to CRSP Lipper Objective and Classification Codes, for Alternative Global Macro Funds, "the strategy is typically based on forecasts and analysis about interest rate trends, the general flow of funds, political changes, government policies, intergovernmental relations, and other broad systemic factors."

sample summary statistics, calculated from net and gross mutual fund returns, following the rolling window methodology from Section 4. To interpret in-sample variation in long-horizon multi-factor timing skills across flexible portfolio allocation mutual fund managers, we provide the mean of each measure within each quartile of in-sample values. Q4 represents the average measure value in the top quartile, and Q1 represents the average value in the bottom quartile.

The top quartile values of all measures are significantly higher than those in the bottom quartile, suggesting significantly better multi-factor timing skills among the top flexible portfolio allocation mutual fund managers compared to the bottom ones.

6.3 Mutual Fund Out-of-Sample Analysis

We now consider whether employing our measures can provide economic value to investors. In other words, we explore whether our measures can identify skilled managers with statistically and economically superior performance out-of-sample. In Table 13, we follow the methodology for out-of-sample tests from Section 4 based on top and bottom quartile portfolios. Unlike simulated timers in the previous section, mutual funds have various life spans. In our analysis, when a fund stops reporting, we assume the fund is defunct and redistribute the remaining capital in the fund equally among surviving portfolio funds. We report attrition rates as average annual rates at which mutual funds stop reporting.

For flexible portfolio allocation mutual funds, the top quartile D_ALPHA portfolio outperforms the bottom quartile D_ALPHA portfolio, generating statistically significant out-of-sample separation alphas based on both net and gross mutual fund returns. Every other measure fails to provide significant separation between top and bottom portfolios out-of-sample. It is notable that the top quartile D_ALPHA portfolio fails to deliver positive out-of-sample net alpha, but delivers positive, yet statistically insignificant, out-of-sample gross alpha, which is consistent with prior mutual fund literature, going back to Carhart (1997).

We conclude that the out-of-sample results for D_ALPHA are consistent with the presence of persistent skill in long-horizon multi-factor timing activity among flexible portfolio allocation mutual fund managers.[35] We further discuss the results in the section that follows.

7 Discussion and Conclusion

In this paper, we consider long-horizon multi-factor timing activity of active portfolio managers based on macroeconomic trends among multiple factors. Building on Jensen (1972), we introduce the new indirect measure of long-horizon timing activity, D_ALPHA, and compare its efficacy with multi-factor extensions of the HM timing measure and conditional COST measures. We demonstrate that D_ALPHA successfully captures long-horizon multi-factor timing in eight- and five-factor long-horizon timing simulations. In simulations, we observe strong positive correlations among D_ALPHA, ALPHA, and COND_ALPHA, provided in Table 14.[36]

In our simulations, we reduce active portfolio management to specifically simulated long-horizon multi-factor timing activity. Our multi-factor timing simulation assumes full investment in one of the considered factors without leverage or derivatives. It is possible for real-life long-horizon multi-factor timing activity to have features not well reflected by our simulation and out-of-sample tests. For example, portfolio managers could be timing other investable factors not

[35] Considering the full sample of mutual funds, D_ALPHA does not exhibit out-of-sample predictive power, which is not surprising, given that the vast majority of equity mutual funds have investment objectives that preclude economically meaningful timing activities. These results are available upon request.

[36] As discussed in Section 4, ALPHA could be interpreted as a timing measure specifically in the context of our simulations, given that timing is the only simulated dimension of active portfolio management. Also, D_ALPHA is the alpha-based measure; hence it is not surprising to see high correlations among D_ALPHA, ALPHA, and COND_ALPHA in timing-only simulations. We confirm that D_ALPHA captures a distinct aspect of long-term multi-factor timing activity, not reflected in ALPHA and COND_ALPHA, in the mutually exclusive portfolio analysis in Tables 4 and 7, further articulating the importance of this analysis in footnote 26.

included in models (15) and (16), and strategically employing leverage and derivatives. Also, our simulation assumes persistent long-horizon multi-factor timing. It does not account for sporadic long-horizon multi-factor timing activity that portfolio managers may undertake infrequently only when clear multi-factor timing opportunities present themselves.[37]

In our analysis of hedge and mutual funds with strategies that are consistent with long-horizon multi-factor timing, macro diversified hedge funds and flexible portfolio allocation mutual funds, we find D_ALPHA out-of-sample results that are consistent with the presence of persistent skill in long-horizon multi-factor timing activity, while all other measures, including ALPHA and COND_ALPHA, do not exhibit significant predictive power. ALPHA and COND_ALPHA are traditionally interpreted measures of all active portfolio management, including activity unrelated to timing, such as security selection, arbitrage, derivative speculation, and risk management, among others. It is plausible to conjecture that macro diversified hedge funds engage in a wide variety of such activities, reflected by ALPHA and COND_ALPHA, but not captured by D_ALPHA, given low correlations of 0.120 between D_ALPHA and ALPHA and 0.021 between D_ALPHA and COND_ALPHA (presented in Table 14). We conjecture that such activities either have dubious out-of-sample success or interfere with multi-factor timing activity captured by D_ALPHA. On the other hand, correlations of 0.363 between D_ALPHA and ALPHA and 0.348 between D_ALPHA and COND_ALPHA in flexible allocation mutual funds are relatively strong and statistically significant. The out-of-sample results for ALPHA and COND_ALPHA in flexible allocation mutual funds are qualitatively similar to the results for D_ALPHA but miss to deliver statistically significant separation between the top and bottom portfolios. This is consistent with

[37] For example, Brunnermeier and Nagel (2004) documented that many hedge fund managers successfully timed into growth technology stocks during the “dot com” bubble in the late 1990s, exiting their positions before the bubble burst in 2000. In “The Big Short” (2010), Michael Lewis provides accounts of the managers who successfully timed the economic meltdown of 2008-2009.

the conjecture that multi-factor timing is likely the most prominent portfolio management activity in flexible allocation mutual funds.

Finally, as mentioned earlier, D_ALPHA is an indirect measure of long-horizon timing activity, reflecting long-horizon timing activity through differences in long- and short-term alphas. Its estimates for individual funds may not be well suited for straightforward quantitative interpretation in-sample, given that its measurement windows are not likely to perfectly match timing horizons of active portfolio managers. Furthermore, D_ALPHA, being the derivative measure of ALPHA, may reflect time-varying profitability from other active portfolio management skills, potentially reflected by ALPHA.[38] However, such time-varying profitability could arguably be considered an indirect form of multi-factor timing.

We conclude that, based on our analysis and subject to the caveats above, out of all considered measures, only the indirect measure, D_ALPHA, consistently captures long-horizon multi-factor activity out-of-sample in simulations and in investment styles that rely on long-horizon multi-factor timing, which are macro diversified in hedge funds and flexible portfolio allocation in mutual funds.

[38] It is possible for active portfolio management skills to generate time-varying profitability patterns depending on macroeconomic conditions. For example, skills in investing in distressed securities, emerging markets, or convertible arbitrage, among others.

References

Agarwal, V., C. Green, and H. Ren. 2018. Alpha or beta in the eye of the beholder: What drives hedge fund flows? *Journal of Financial Economics* 127: 417-434.

Agarwal, V., and N. Y. Naik. 2000. On taking the alternative route: risks, rewards, and performance persistence of hedge funds. *Journal of Alternative Investments* 2: 6-23.

Agarwal, V., and N. Y. Naik. 2004. Risk and portfolio decisions involving hedge funds. *Review of Financial Studies* 17: 63-98.

Amihud, Y., and R. Goyenko. 2013. Mutual fund's R^2 as predictor of performance. *Review of Financial Studies* 26: 667-694.

Ang, A. 2014. Asset management: A systematic approach to factor investing. *Oxford University Press*.

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

Aragon, G. 2002. Timing multiple markets: Theory and evidence. *SSRN manuscript*: https://ssrn.com/abstract=415740.

Arnott, R., C. Harvey, V. Kalesnik, and J. Linnainmaa. 2019. Alice's adventures in factorland: Three blunders that plague factor investing. *Journal of Portfolio Management* 45: 18-36.

Asness, C., R. J. Krail, and J. M. Liew. 2001. Do hedge funds hedge? *Journal of Portfolio Management* 28: 6-19.

Avramov, D., R. Kosowski, N. Y. Naik, and M. Teo. 2011. Hedge funds, managerial skill, and macroeconomic variables. *Journal of Financial Economics* 99: 672-692.

Avramov, D., and R. Wermers. 2006. Investing in mutual funds when returns are predictable. *Journal of Financial Economics* 81: 339-377.

Bali, T. G., S. J. Brown, and M. O. Caglayan. 2011. Do hedge funds' exposures to risk factors predict their future returns? *Journal of Financial Economics* 101: 36-68.

Bali, T. G., S. J. Brown, and M. O. Caglayan. 2012. Systematic risk and the cross section of hedge fund returns. *Journal of Financial Economics* 106: 114-131.

Bali, T. G., S. J. Brown, and M. O. Caglayan. 2014. Macroeconomic risk and hedge fund returns. *Journal of Financial Economics* 114: 1-19.

Barberis, N. 2000. Investing for the long run when returns are predictable. *Journal of Finance* 55: 225-264.

Ben-David, I., F. Franzoni, B. Kim, and R. Moussawi. 2023. Competition for attention in the ETF space. *Review of Financial Studies* 36: 987-1042.

Ben-Rephael, A., S. Kandel, and A. Wohl. 2012. Measuring investor sentiment with mutual fund flows. *Journal of Financial Economics* 104: 363-382.

Bergstresser, D., J. M. R. Chalmers, and P.Tufano. 2009. Assessing the costs and benefits of brokers in the mutual fund industry. *Review of Financial Studies* 22: 4129-4156.

Bodnaruk, A., B. Chokaev, and A. Simonov. 2018. Downside risk timing by mutual funds. *Review of Asset Pricing Studies* 9: 171-196.

Boguth, O., M. Carlson, A. Fisher, and M. Simutin. 2011. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics* 102: 363-389.

Bollen, N. P. B., and J. A. Busse. 2001. On the timing ability of mutual fund managers. *Journal of Finance* 56: 1075-1094.

Bollen, N. P. B., and J. A. Busse. 2004. Short-term persistence in mutual fund performance. *Review of Financial Studies* 18: 569-597.

Bollen, N. P. B., and R. E. Whaley. 2009. Hedge fund risk dynamics: Implications for performance appraisal. *Journal of Finance* 64: 985-1035.

Brunnermeier, M. K., and S. Nagel. 2004. Hedge funds and the technology bubble. *Journal of Finance* 59: 2013-2040.

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

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

Cai, L., and B. Liang. 2012a. Asset allocation dynamics in the hedge fund industry. *Journal of Investment Management* 10: 35-59.

Cai, L., and B. Liang. 2012b. On the dynamics of hedge fund strategies. *Journal of Alternative Investments* 14: 51-68.

Cao, C., Y. Chen, B. Liang, and A. W. Lo. 2013. Can hedge funds time market liquidity? *Journal of Financial Economics* 109: 493-516.

Cederburg, S., M.S. O'Doherty, N.E. Savin, and A. Tiwari. 2018. Conditional benchmarks and predictors of mutual fund performance. *Critical Finance Review* 7: 331-372.

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

Chance, D. M., and M. L. Hemler. 2001. The performance of professional market timers: Daily evidence from executed strategies. *Journal of Financial Economics* 62: 377-411.

Chang, E. C., and W. G. Lewellen. 1984. Market timing and mutual fund investment performance. *Journal of Business* 57: 57-72.

Chen, Y. 2007. Timing ability in the focus market of hedge funds. *Journal of Investment Management* 5: 66-98.

Chen, Y., W. Ferson, and H. Peters. 2010. Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics* 98: 72-89.

Chen, Y., and B. Liang. 2007. Do market timing hedge funds time the market? *Journal of Financial and Quantitative Analysis* 42: 827-856.

Chernov, M., L. A. Lochstoer, and S. R. H. Lundeby. 2022. Conditional dynamics and the multihorizon risk-return trade-off. *Review of Financial Studies* 35: 1310-1347.

Chuprinin, O., and D. Sosyura. 2018. Family descent as a signal of managerial quality: Evidence from mutual funds. *Review of Financial Studies* 31: 3756-3820.

Comer, G. 2006. Hybrid mutual funds and market timing performance. *Journal of Business* 79: 771-797.

Cremers, M., and A. Petajisto. 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22: 3329-3365.

Cremers, M., A. Petajisto, and E. Zitzewitz. 2013. Should benchmark indices have alpha? Revisiting performance evaluation. *Critical Finance Review* 2: 1-48.

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

Duanmu, J., A. Malakhov, and W. R. McCumber. 2018. Beta active hedge fund management. *Journal of Financial and Quantitative Analysis* 53: 2525-2558.

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

Ehsani, S., and J. T. Linnainmaa. 2022. Factor momentum and the momentum factor. *Journal of Finance* 77: 1877-1919.

Elton, E., M. Gruber, and C. Blake. 2012. An examination of mutual fund timing ability using monthly holdings data. *Review of Finance* 16: 619-645.

Evans, R. B. 2010. Mutual fund incubation. *Journal of Finance* 65: 1581-1611.

Fama, E. 1972. Components of investment performance. *Journal of Finance* 27: 551-567.

Fama, E., and K. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3-56.

Fama, E., and K. French. 2010. Luck versus skill in the cross-section of mutual fund performance. *Journal of Finance* 65: 1915-1947.

Ferson, W., and R. W. Schadt. 1996. Measuring fund strategy and performance in changing economic conditions. *Journal of Finance* 51: 425-460.

Fulkerson, J. 2013. Is timing everything? The value of mutual fund manager trades. *Financial Management* 42: 243-261.

Fung, H.-G., X. E. Xu, and J. Yau. 2002. Global hedge funds: Risk, return, and market timing. *Financial Analysts Journal* 58: 19-30.

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

Fung, W., and D. A. Hsieh. 2001. The risk in hedge fund strategies: Theory and evidence from trend followers. *Review of Financial Studies* 14: 313-341.

Fung, W., and D. A. Hsieh. 2004. Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal* 60: 65-80.

Fung, W., D. A. Hsieh, N. Y. Naik, and T. Ramadorai. 2008. Hedge funds: Performance, risk, and capital formation. *Journal of Finance* 63: 1777-1803.

Glushkov, D. 2016. How smart are smart beta exchange-traded funds? Analysis of relative performance and factor exposure. *Journal of Investment Consulting* 17: 50-74.

Goetzmann, W., J. Ingersoll, and Z. Ivkovic. 2000. Monthly measurement of daily timers. *Journal of Financial and Quantitative Analysis* 35: 257-290.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch. 2007. Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies* 20: 1503-1546.

Graham, J. R., and C. R. Harvey. 1996. Market timing ability and volatility implied in investment newsletters' asset allocation recommendations. *Journal of Financial Economics* 42: 397-422.

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

Gupta, T., and B. Kelly. 2019. Factor momentum everywhere. *Journal of Portfolio Management* 45: 13-36.

Haddad, V., S. Kozak, and S. Santosh. 2020. Factor timing. *Review of Financial Studies* 33: 1980-2018.

Handa, P., S. P. Kothari, and C. Wasley. 1989. The relation between the return interval and betas: Implication for the size effect. *Journal of Financial Economics* 23: 79-100.

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

Henriksson, R. D., and R. C. Merton. 1981. On market timing and investment performance II: Statistical procedures for evaluating forecasting skills. *Journal of Business* 54: 513-534.

Hunter, D., E. Kandel, S. Kandel, and R. Wermers. 2014. Mutual fund performance evaluation with active peer benchmarks. *Journal of Financial Economics* 112: 1-29.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48: 65-91.

Jagannathan, R., and R. A. Korajczyk. 1986. Assessing the market timing performance of managed portfolios. *Journal of Business* 59: 217-235.

Jagannathan, R., A. Malakhov, and D. Novikov. 2010. Do hot hands exist among hedge fund managers? An empirical evaluation. *Journal of Finance* 65: 217-255.

Jagannathan, R., A. Malakhov, and D. Novikov. 2013. Correcting for the backfill bias in hedge fund databases, Appendix to: Do hot hands exist among hedge fund managers? An empirical evaluation. *SSRN manuscript*: https://ssrn.com/abstract=2278686.

Jensen, M. C. 1968. The performance of mutual funds in the period 1945-1964. *Journal of Finance* 23: 389-416.

Jensen, M. C. 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* 42: 167–247.

Jensen, M. C. 1972. Optimal utilization of market forecasts and the evaluation of investment performance. In G. P. Szego and K. Shell (eds.), *Mathematical methods in investment and finance*. Amsterdam: North-Holland.

Jiang, G., T. Yao, and T. Yu. 2007. Do mutual funds time the market? Evidence from portfolio holdings. *Journal of Financial Economics* 86: 724-758.

Jiang, W. 2003. A nonparametric test of market timing. *Journal of Empirical Finance* 10: 399-425.

Kacperczyk, M., C. Sialm, and L. Zheng. 2008. Unobserved actions of mutual funds. *Review of Financial Studies* 21: 2379-2416.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp. 2014. Time-varying fund manager skill. *Journal of Finance* 69: 1455-1484.

Kamara, A., R. A. Korajczyk, X. Lou, and R. Sadka. 2016. Horizon pricing. *Journal of Financial and Quantitative Analysis* 51: 1769-1793.

Kosowski, R., A. Timmermann, R. Wermers, and H. White. 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance* 61: 2551-2595.

Lee, C. F., C. Wu, and K. C. J. Wei. 1990. The heterogeneous investment horizon and the capital asset pricing model: Theory and implications. *Journal of Financial and Quantitative Analysis* 25: 361-376.

Lehmann, B., and D. Modest. 1987. Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance* 42: 233-265.

Levhari, D., and H. Levy. 1977. The capital asset pricing model and the investment horizon. *Review of Economics and Statistics* 59: 92-104.

Lewellen, J., and S. Nagel. 2006. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82: 289–314.

Lewis, M. 2010. The big short: Inside the doomsday machine. *W. W. Norton & Company*.

Malkiel, B. G. 1995. Returns from investing in equity mutual funds 1971 to 1991. *Journal of Finance* 50: 549-572.

Mamaysky, H., M. Spiegel, and H. Zhang. 2008. Estimating the dynamics of mutual fund alphas and betas. *Review of Financial Studies* 21: 233-264.

Mateus, C., I. Mateus, and M. Soggiu. 2020. Do smart beta ETFs deliver persistent performance? *Journal of Asset Management* 21: 413-427.

Merton, R. C. 1981. On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *Journal of Business* 54: 363-406.

Moskowitz, T. J. 2000. Discussion: Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *Journal of Finance* 55: 1695-1703.

Namvar, E., B. Phillips, K. Pukthuanthong, and P. R. Rau. 2016. Do hedge funds dynamically manage systematic risk? *Journal of Banking and Finance* 64: 1-15.

Patton, A. J., and T. Ramadorai. 2013. On the high-frequency dynamics of hedge fund risk exposures. *Journal of Finance* 68: 597-635.

Sharpe, W. F. 1975. Likely gains from market timing. *Financial Analysts Journal* 31: 60-69.

Sharpe, W. F. 1992. Asset allocation: Management style and performance management. *Journal of Portfolio Management* 18: 7–19.

Titman, S., and C. Tiu. 2011. Do the best hedge funds hedge? *Review of Financial Studies* 24: 123-168.

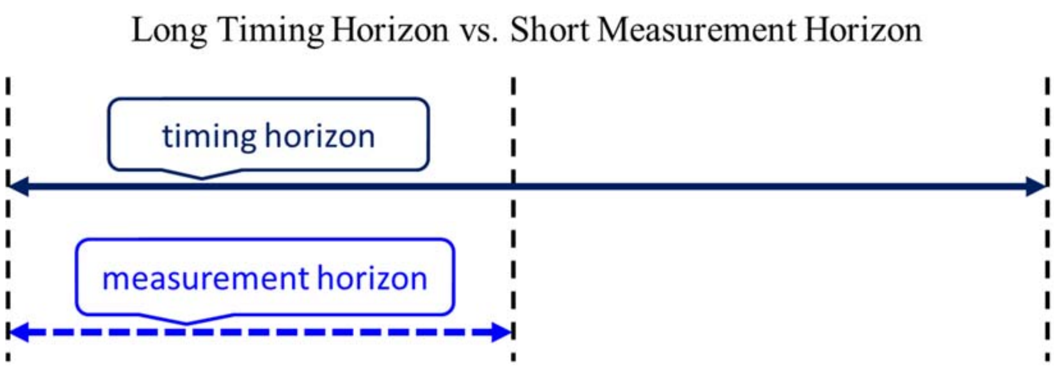
Treynor, J., and K. Mazuy. 1966. Can mutual funds outguess the market? *Harvard Business Review* 44: 131-136.

Wermers, R. 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *Journal of Finance* 55: 1655-1695.

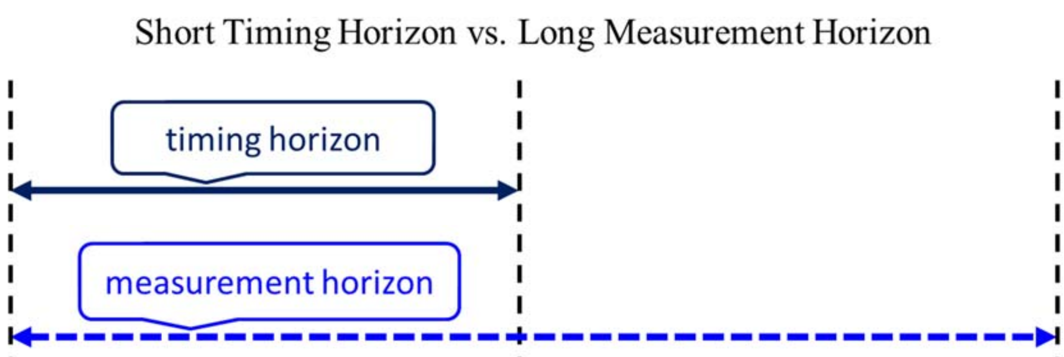
Figure 1: Timing horizon vs. measurement horizon, based on Jensen (1972)

Figure 1A displays a long timing horizon relative to a short measurement horizon, with portfolio manager's performance evaluated according to the single factor model from Jensen (1972): $r_t - r_t^f = \alpha + \beta\left(r_t^{market} - r_t^f\right) + \varepsilon_t$. α reflects security selection activity, but does not reflect timing activity.

Figure 1B displays a short timing horizon relative to a long measurement horizon, with portfolio manager's performance evaluated according to the single factor model from Jensen (1972): $r_t - r_t^f = \alpha + \beta\left(r_t^{market} - r_t^f\right) + \varepsilon_t$. α reflects both security selection activity and timing activity together.



A: α reflects security selection, α <u>does not</u> reflect timing activity

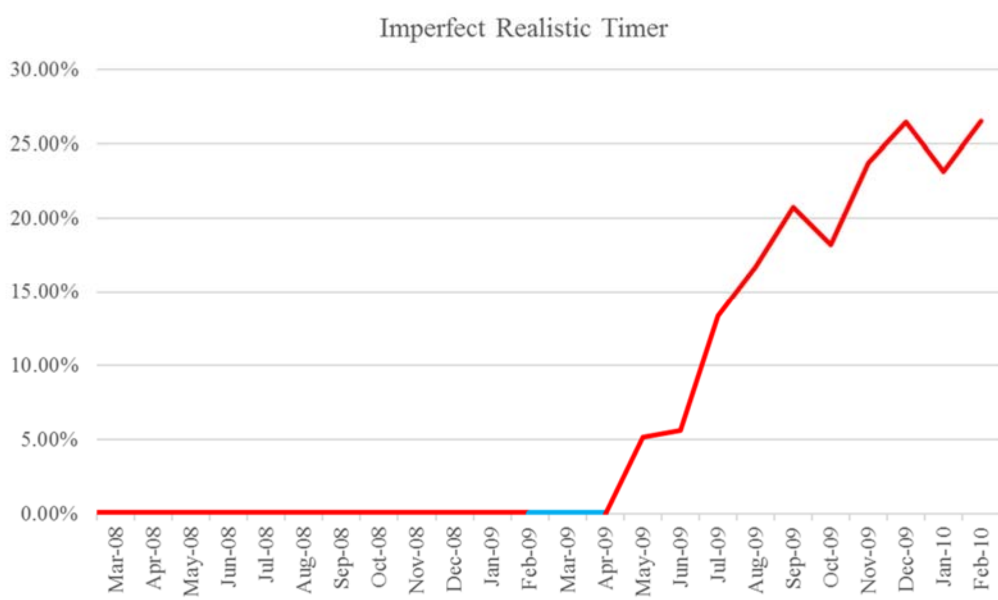


B: α reflects both security selection <u>and</u> timing activity together

Figure 2: Market timing of an imperfect realistic long-term timer

Figure 2A displays cumulative market excess return and an imperfect realistic long-term market timer from March 2008 to February 2010. The blue line represents the timer investing in the U.S. Treasury Bills, and the red line represents the timer investing in the S&P 500 index fund. Figure 2B illustrates the corresponding return payoff of the imperfect realistic timer.

A



B

Figure 3: Motivation for the proposed indirect long-term timing measure, D_ALPHA

Figure 3 displays the return payoff of the imperfect realistic long-term timer from March 2008 to February 2010 in the motivating example. The text box presents the imperfect realistic timer's alphas from the single factor model, $r_t - r_t^f = \alpha + \beta\left(r_t^{market} - r_t^f\right) + \varepsilon_t$, evaluated over the overall 24-month window and three rolling 12-month windows. The detailed regression results are presented in Table 1.

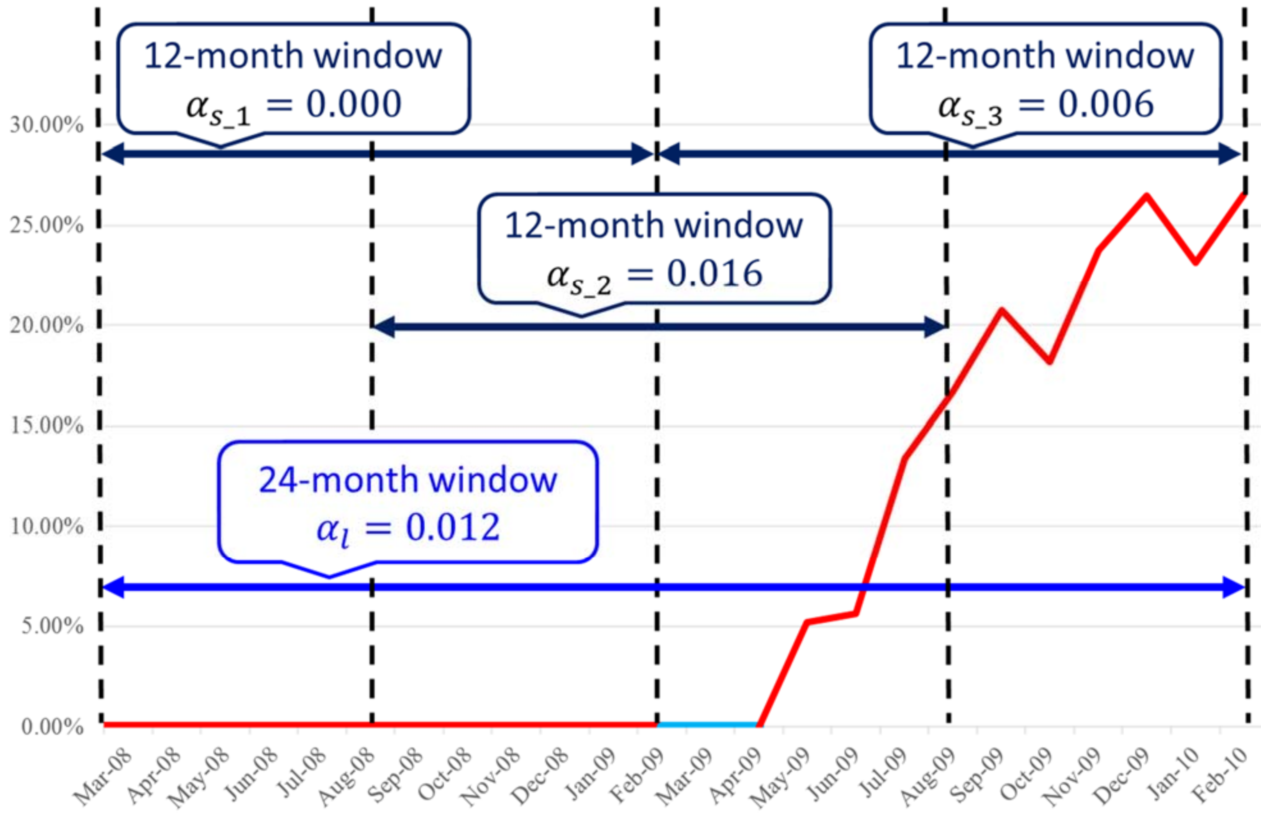


Figure 4: Hedge fund factors from model (15) and the optimal timing pattern in the eight-factor simulation

Figure 4 displays the optimal multi-factor timing pattern for Fung and Hsieh (2004) eight factors (15) and the 30-day U.S. Treasury Bill return, along with the evolution of natural log portfolio values from investing $1 on January 1, 1994. The thick solid line represents the optimal timing pattern value, and dotted lines represent log portfolio values if invested in suboptimal factors at each switch point. The sample period is from January 1994 to December 2018.

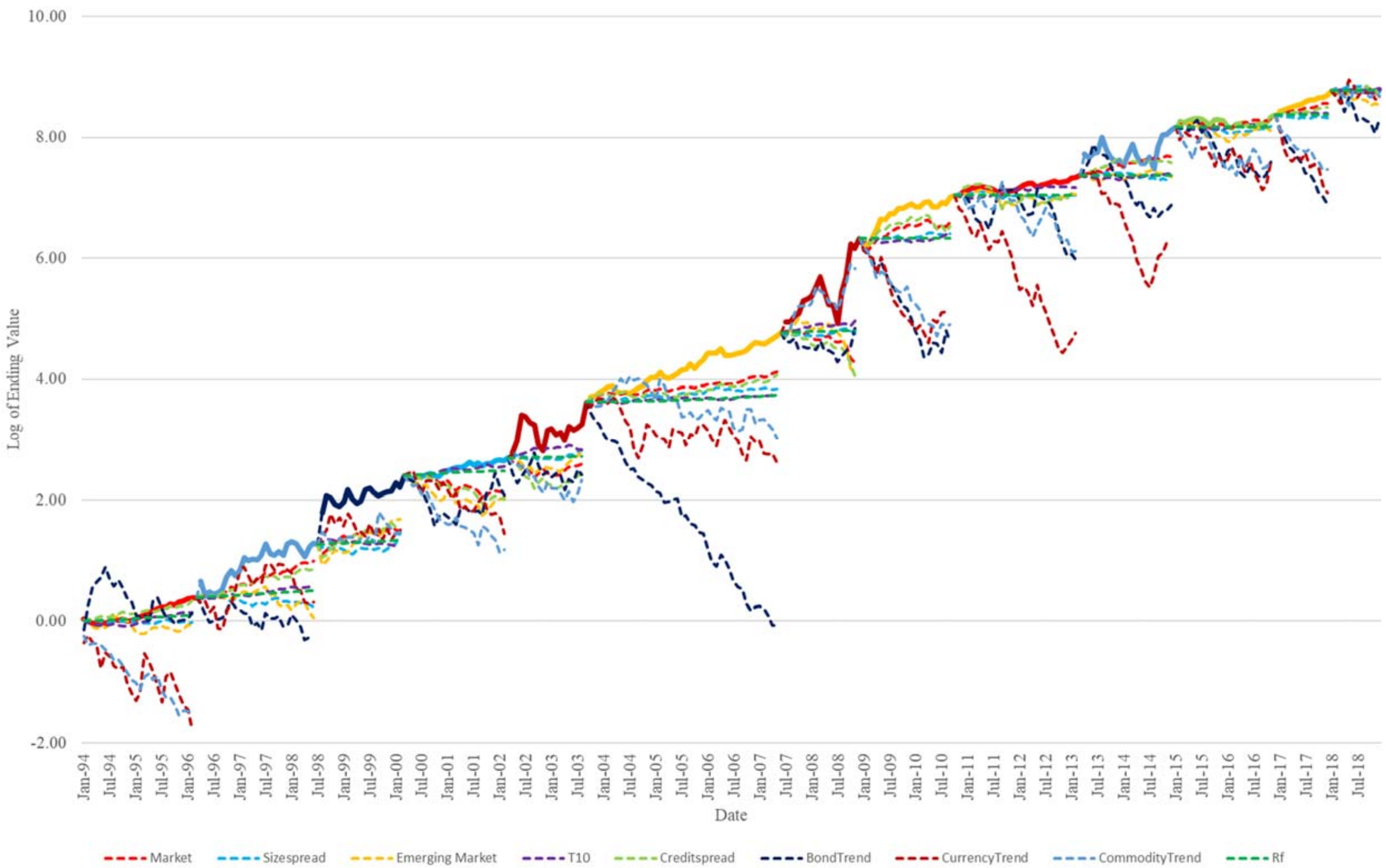


Figure 5: Mutual fund factors from model (16) and the optimal timing pattern in the five-factor simulation

Figure 5 displays the optimal multi-factor timing pattern for five factors from the mutual fund performance evaluation model (16) and the 30-day U.S. Treasury Bill return, along with the evolution of natural log portfolio values from investing $1 on January 1, 1994. The thick solid line represents the optimal timing pattern value, and dotted lines represent log portfolio values if invested in suboptimal factors at each switch point. The sample period is from January 1994 to December 2018.

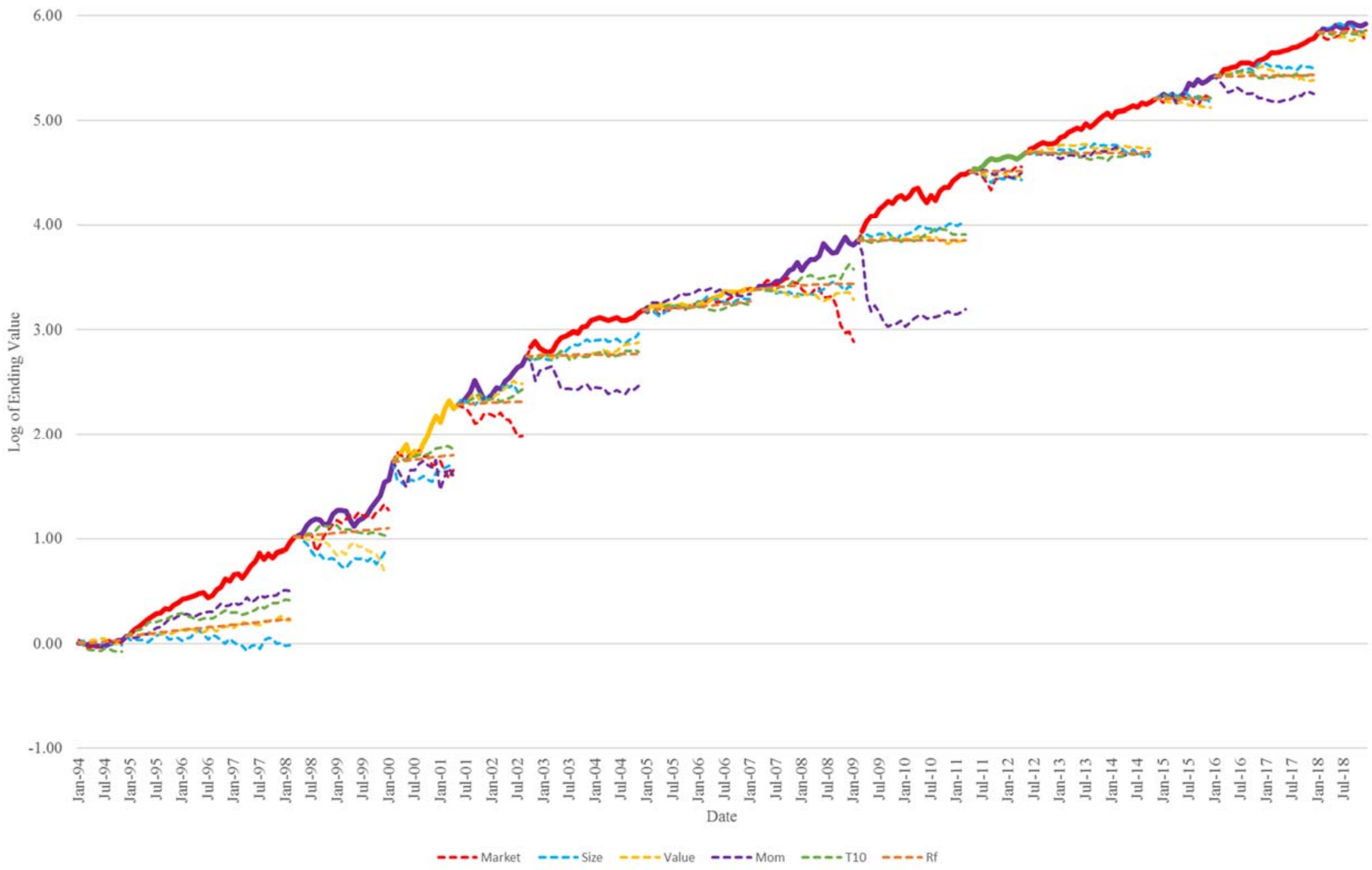


Table 1: Single factor model results for the motivating example

Table 1 presents the single-factor model results for the imperfect realistic timer in the motivating example, displayed in Figures 3 and 4. The imperfect realistic timer invests in the U.S. Treasury Bills from March 2008 to April 2009, and in the S&P 500 index fund from May 2009 to February 2010. The timer performance is evaluated with the model, $r_t - r_t^f = \alpha + \beta\left(r_t^{market} - r_t^f\right) + \varepsilon_t$. Panel A presents results over the 24-month window, from March 2008 to February 2010. Panel B presents results over the first 12-month window, from March 2008 to February 2009. Missing values for t-statistics and P-values are due to the dependent variable in the single factor regression being zero for the entire 12-month window from March 2008 to February 2009. Panel C presents results over the second 12-month window, from September 2008 to August 2009. Panel D presents results over the third 12-month window, from March 2009 to February 2010.

Panel A: ***Results for the 24-month Window, March 2008 - February 2010***

	Coefficient	t-stat	P-value
α	0.012	2.433	0.024
R_{market} - R_f	0.177	2.491	0.021

Panel B: ***Results for the First 12-month Window, March 2008 - February 2009***

	Coefficient	t-stat	P-value
α	0	.	.
R_{market} - R_f	0	.	.

Panel C: ***Results for the Second 12-month Window, September 2008 - August 2009***

	Coefficient	t-stat	P-value
α	0.016	2.207	0.052
R_{market} - R_f	0.135	1.656	0.129

Panel D: ***Results for the Third 12-month Window, March 2009 - February 2010***

	Coefficient	t-stat	P-value
α	0.006	0.501	0.627
R_{market} - R_f	0.424	1.970	0.077

Table 2: Summary statistics from the eight-factor timing simulation

Table 2 presents summary statistics of proposed timing measure values for the nine types of imperfect timers. All measures are based on the Fung and Hsieh (2004) eight-factor model (15). The imperfect timers make errors in both timing of switching into the optimal factor and in the optimal factor selection. We simulate three levels of accuracy in both dimensions, which yields nine possible combinations in timing accuracy combined with factor selection accuracy: we simulate 10,000 timers for each combination. We simulate timing accuracy by allowing the factor switch time within two, four, and six months prior to the optimal switch time (the optimal switch time included) via uniform distribution. We simulate factor selection accuracy in that the optimal factor at each switch point is selected with probabilities of 90%, 60%, and 30%. For each timer, we calculate each measure over four-year rolling windows over the sample period of 25 years from January 1994 to December 2018. This yields 21,000 (21*10,000 timers) observations for each category of accuracy for the unconditional measures, D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, and 19,000 (19*10,000 timers) observations for conditional measures, COND_MAX_FT, COND_TOTAL_FT, COND_VOL, COND_ALPHA, given that calculating the conditional measures requires 24-month history data to calculate past 24-month beta exposures. Standard deviations are presented in the Appendix.

Timing Accuracy		**Factor Selection Accuracy**		
		90%	**60%**	**30%**
2 months	D_ALPHA	0.818	0.509	0.195
	MAX_ALL	0.188	0.117	0.046
	MAX_SINGLE	2.405	2.169	1.940
	SUM_ALL	0.211	0.138	0.040
	ALPHA	2.220	1.379	0.536
	COND_MAX_FT	0.863	0.850	0.793
	COND_TOTAL_FT	0.580	0.377	0.146
	COND_VOL	0.121	0.067	0.029
	COND_ALPHA	1.437	0.884	0.341
4 months	D_ALPHA	0.772	0.479	0.185
	MAX_ALL	0.157	0.097	0.039
	MAX_SINGLE	2.345	2.108	1.901
	SUM_ALL	-0.118	-0.072	-0.028
	ALPHA	1.815	1.127	0.442
	COND_MAX_FT	0.922	0.859	0.973
	COND_TOTAL_FT	0.414	0.303	0.157
	COND_VOL	0.140	0.069	-0.002
	COND_ALPHA	1.142	0.682	0.260
6 months	D_ALPHA	0.711	0.442	0.171
	MAX_ALL	0.130	0.081	0.031
	MAX_SINGLE	2.250	2.028	1.852
	SUM_ALL	-0.369	-0.220	-0.084
	ALPHA	1.400	0.873	0.336
	COND_MAX_FT	1.073	0.852	1.354
	COND_TOTAL_FT	0.411	0.302	0.097
	COND_VOL	0.154	0.070	0.052
	COND_ALPHA	0.682	0.408	0.151

Table 3: Out-of-sample performance in the eight-factor timing simulation

Table 3 reports out-of-sample performance of portfolios based on top and bottom deciles of past values of each measure. Each portfolio is rebalanced annually based on the updated ranking with respect to preceding four-year window measure values. For D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, portfolios are initially formed on December 31, 1997. For COND_MAX_FT, COND_TOTAL_FT, COND_VOL, COND_ALPHA, portfolios are initially formed on December 31, 1999. We invest $1 in each timer within a portfolio at the beginning of the year and track its performance until December 31, 2018. The top-bottom portfolio is equivalent to taking a long position in the top decile portfolio and a short position in the bottom decile portfolio. The reported alpha and its corresponding T-value are from the Fung and Hsieh (2004) eight-factor model (15). Return is the mean excess return. The End Value is the dollar value of the initial $1 investment as of December 2018.

Variable	Portfolio	Alpha	T-Alpha	Return	Sharpe	End Value
D_ALPHA	Top	2.122	4.429	1.542	0.167	26.774
	Bottom	0.449	1.625	0.081	0.014	1.232
	Top-Bottom	1.673	3.650	1.461	0.197	20.699
MAX_ALL	Top	1.322	3.562	0.930	0.130	8.280
	Bottom	1.146	3.562	0.791	0.123	6.592
	Top-Bottom	0.176	0.515	0.139	0.026	1.004
MAX_SINGLE	Top	1.412	3.532	1.056	0.131	9.877
	Bottom	0.965	3.688	0.790	0.160	8.025
	Top-Bottom	0.447	1.631	0.266	0.053	1.458
SUM_ALL	Top	1.424	3.352	0.842	0.100	5.475
	Bottom	1.173	3.313	1.020	0.153	11.323
	Top-Bottom	0.251	0.518	-0.178	-0.022	0.308
ALPHA	Top	2.382	4.823	1.851	0.201	57.585
	Bottom	0.519	2.013	0.092	0.016	1.262
	Top-Bottom	1.863	4.741	1.759	0.285	51.837
COND_MAX_FT	Top	1.327	3.525	1.084	0.145	8.770
	Bottom	0.749	3.091	0.688	0.146	5.042
	Top-Bottom	0.577	2.629	0.396	0.100	2.075
COND_TOTAL_FT	Top	1.501	3.567	1.257	0.153	11.535
	Bottom	1.154	3.940	0.897	0.151	7.050
	Top-Bottom	0.346	1.473	0.360	0.098	1.952
COND_VOL	Top	1.832	3.883	1.550	0.182	21.542
	Bottom	0.713	2.286	0.493	0.078	2.692
	Top-Bottom	1.120	2.792	1.057	0.178	7.619
COND_ALPHA	Top	1.588	4.252	1.468	0.210	22.147
	Bottom	0.892	2.556	0.520	0.075	2.647
	Top-Bottom	0.696	2.926	0.949	0.265	7.466

Table 4: Mutually exclusive portfolio performance in the eight-factor timing simulation

Table 4 presents out-of-sample performance of mutually exclusive portfolios. We first form top decile portfolios sorted by one measure and then exclude any timers that are also in the top decile of the other measure. For example, D_ALPHA ex ALPHA is the portfolio consisting of timers present in the top decile of the D_ALPHA-sorted portfolio but not in the top decile of the ALPHA-sorted portfolio. We only consider measures that demonstrate predictive power in the out-of-sample test in Table 3. Portfolios based on unconditional measures, D_ALPHA and ALPHA, are formed on December 31, 1997. All portfolios that involve conditional measures, COND_MAX_FT, COND_VOL, COND_ALPHA, are formed on December 31, 1999. The reported alpha and its corresponding T-value are from the Fung and Hsieh (2004) eight-factor model (15). Return is the mean excess return. The End Value is the dollar value of the initial $1 investment as of December 2018.

Variable	Mutually Exclusive Portfolio	Alpha	T-Alpha	Return	Sharpe	End Value
D_ALPHA	D_ALPHA ex ALPHA	1.718	4.277	1.143	0.161	12.717
	D_ALPHA ex COND_MAX_FT	1.761	3.691	1.480	0.168	17.448
	D_ALPHA ex COND_VOL	1.711	3.697	1.415	0.163	15.482
	D_ALPHA ex COND_ALPHA	1.760	3.707	1.427	0.161	15.392
ALPHA	ALPHA ex D_ALPHA	2.135	4.787	1.672	0.215	60.092
	ALPHA ex COND_MAX_FT	1.967	4.070	1.813	0.209	37.423
	ALPHA ex COND_VOL	1.907	4.100	1.753	0.208	34.191
	ALPHA ex COND_ALPHA	2.035	4.000	1.824	0.198	35.397
COND_MAX_FT	COND_MAX_FT ex D_ALPHA	1.286	3.407	1.080	0.147	8.831
	COND_MAX_FT ex ALPHA	1.299	3.450	1.080	0.147	8.811
	COND_MAX_FT ex COND_VOL	1.249	3.359	1.001	0.134	7.258
	COND_MAX_FT ex COND_ALPHA	1.321	3.486	1.073	0.143	8.497
COND_VOL	COND_VOL ex D_ALPHA	1.825	3.886	1.575	0.188	23.331
	COND_VOL ex ALPHA	1.807	3.863	1.539	0.184	21.570
	COND_VOL ex COND_MAX_FT	1.883	3.863	1.609	0.186	24.090
	COND_VOL_FT ex COND_ALPHA	1.822	3.862	1.552	0.183	21.780
COND_ALPHA	COND_ALPHA ex D_ALPHA	1.447	4.155	1.437	0.224	22.311
	COND_ALPHA ex ALPHA	1.338	4.061	1.285	0.207	16.287
	COND_ALPHA ex COND_MAX_FT	1.590	4.262	1.473	0.211	22.444
	COND_ALPHA ex COND_VOL	1.576	4.235	1.476	0.213	22.759

Table 5: Summary statistics from the five-factor timing simulation

Table 5 presents summary statistics of proposed timing measure values for the nine types of imperfect timers. All measures are based on the five-factor model (16). The imperfect timers make errors in both timing of switching into the optimal factor and in the optimal factor selection. We simulate three levels of accuracy in both dimensions, which yields nine possible combinations in timing accuracy combined with factor selection accuracy: we simulate 10,000 timers for each combination. We simulate timing accuracy by allowing the factor switch time within two, four, and six months prior to the optimal switch time (the optimal switch time included) via uniform distribution. We simulate factor selection accuracy in that the optimal factor at each switch point is selected with probabilities of 90%, 60%, and 30%. For each timer, we calculate each measure over four-year rolling windows over the sample period of 25 years from January 1994 to December 2018. This yields 21,000 (21*10,000 timers) observations for each category of accuracy for the unconditional measures, D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, and 19,000 (19*10,000 timers) observations for conditional measures, COND_MAX_FT, COND_TOTAL_FT, COND_VOL, and COND_ALPHA, given that calculating the conditional measures requires 24-month history data to calculate past 24-month beta exposures. Standard deviations are presented in the Appendix.

Timing Accuracy		**Factor Selection Accuracy**		
		90%	**60%**	**30%**
2 months	D_ALPHA	0.337	0.199	0.061
	MAX_ALL	0.224	0.132	0.041
	MAX_SINGLE	0.849	0.729	0.625
	SUM_ALL	0.207	0.126	0.033
	ALPHA	0.876	0.518	0.157
	COND_MAX_FT	0.293	0.206	0.135
	COND_TOTAL_FT	0.308	0.172	0.046
	COND_VOL	-0.084	-0.044	-0.012
	COND_ALPHA	0.715	0.427	0.135
4 months	D_ALPHA	0.315	0.186	0.057
	MAX_ALL	0.192	0.114	0.035
	MAX_SINGLE	0.811	0.710	0.621
	SUM_ALL	0.019	0.014	-0.002
	ALPHA	0.744	0.441	0.132
	COND_MAX_FT	0.296	0.211	0.137
	COND_TOTAL_FT	0.321	0.179	0.048
	COND_VOL	-0.072	-0.038	-0.009
	COND_ALPHA	0.550	0.333	0.103
6 months	D_ALPHA	0.275	0.162	0.050
	MAX_ALL	0.153	0.091	0.028
	MAX_SINGLE	0.801	0.697	0.609
	SUM_ALL	-0.081	-0.046	-0.016
	ALPHA	0.571	0.338	0.103
	COND_MAX_FT	0.277	0.203	0.135
	COND_TOTAL_FT	0.306	0.167	0.045
	COND_VOL	-0.055	-0.024	-0.005
	COND_ALPHA	0.357	0.216	0.069

Table 6: Out-of-sample performance in the five-factor timing simulation

Table 6 reports out-of-sample performance of portfolios based on top and bottom deciles of past values of each measure. Each portfolio is rebalanced annually based on the updated ranking with respect to preceding four-year window measure values. For D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, portfolios are initially formed on December 31, 1997. For COND_MAX_FT, COND_TOTAL_FT, COND_VOL, COND_ALPHA, portfolios are initially formed on December 31, 1999. We invest $1 in each timer within a portfolio at the beginning of the year and track its performance until December 31, 2018. The top-bottom portfolio is equivalent to taking a long position in the top decile portfolio and a short position in the bottom decile portfolio. The reported alpha and its corresponding T-value are from the five-factor model (16). Return is the mean excess return. The End Value is the dollar value of the initial $1 investment as of December 2018.

Variable	Portfolio	Alpha	T-Alpha	Return	Sharpe	End Value
	Top	0.708	4.452	0.929	0.308	13.601
D_ALPHA	Bottom	0.320	3.911	0.543	0.321	5.593
	Top-Bottom	0.388	2.506	0.386	0.148	2.431
	Top	0.740	4.625	0.984	0.338	15.725
MAX_ALL	Bottom	0.453	3.616	0.646	0.225	6.774
	Top-Bottom	0.286	1.703	0.338	0.109	2.080
	Top	0.499	4.012	0.755	0.314	9.172
MAX_SINGLE	Bottom	0.484	4.946	0.695	0.340	8.057
	Top-Bottom	0.016	0.140	0.060	0.031	1.109
	Top	0.409	3.644	0.639	0.290	6.947
SUM_ALL	Bottom	0.508	3.686	0.733	0.235	8.251
	Top-Bottom	-0.099	-0.650	-0.093	-0.029	0.693
	Top	0.905	5.123	1.148	0.356	23.055
ALPHA	Bottom	0.255	3.315	0.473	0.277	4.691
	Top-Bottom	0.650	3.977	0.675	0.253	4.988
	Top	0.833	4.973	1.068	0.356	13.752
COND_MAX_FT	Bottom	0.351	4.569	0.574	0.360	4.837
	Top-Bottom	0.482	3.391	0.494	0.217	2.906
	Top	0.810	4.857	1.048	0.350	13.146
COND_TOTAL_FT	Bottom	0.347	3.675	0.595	0.307	5.002
	Top-Bottom	0.464	2.908	0.453	0.171	2.595
	Top	0.583	4.885	0.829	0.366	8.363
COND_VOL	Bottom	0.651	4.777	0.905	0.341	9.725
	Top-Bottom	-0.068	-0.588	-0.076	-0.040	0.805
	Top	0.890	4.877	1.138	0.352	15.839
COND_ALPHA	Bottom	0.314	4.184	0.571	0.329	4.777
	Top-Bottom	0.576	3.761	0.567	0.243	3.420

Table 7: Mutually exclusive portfolio performance in the five-factor timing simulation

Table 7 presents out-of-sample performance of mutually exclusive portfolios. We first form top decile portfolios sorted by one measure and then exclude any timers that are also in the top decile of the other measure. For example, D_ALPHA ex ALPHA is the portfolio consisting of timers present in the top decile of the D_ALPHA-sorted portfolio but not in the top decile of the ALPHA-sorted portfolio. We only consider measures that demonstrate predictive power in the out-of-sample test in Table 6. Portfolios based on unconditional measures, D_ALPHA, MAX_ALL, and ALPHA are formed on December 31, 1997. All portfolios that involve conditional measures, COND_MAX_FT, COND_TOTAL_FT, COND_ALPHA, are formed on December 31, 1999. The reported alpha and its corresponding T-value are from the five-factor model (16). Return is the mean excess return. The End Value is the dollar value of the initial $1 investment as of December 2018.

Variable	Mutually Exclusive Portfolio	Alpha	T-Alpha	Return	Sharpe	End Value
	D_ALPHA ex MAX_ALL	0.635	4.091	0.863	0.290	11.576
	D_ALPHA ex ALPHA	0.625	4.294	0.830	0.293	10.755
D_ALPHA	D_ALPHA ex COND_MAX_FT	0.741	4.404	0.993	0.323	11.556
	D_ALPHA ex COND_TOTAL_FT	0.745	4.505	1.001	0.329	11.779
	D_ALPHA ex COND_ALPHA	0.712	4.407	0.944	0.319	10.423
	MAX_ALL ex D_ALPHA	0.672	4.438	0.919	0.333	13.488
	MAX_ALL ex ALPHA	0.631	4.526	0.873	0.339	12.177
MAX_ALL	MAX_ALL ex COND_MAX_FT	0.694	4.181	0.951	0.327	10.609
	MAX_ALL ex COND_TOTAL_FT	0.667	4.107	0.932	0.327	10.211
	MAX_ALL ex COND_ALPHA	0.674	4.317	0.902	0.328	9.596
	ALPHA ex D_ALPHA	0.892	4.956	1.124	0.348	21.776
	ALPHA ex MAX_ALL	0.725	4.090	0.969	0.298	14.773
ALPHA	ALPHA ex COND_MAX_FT	0.883	4.736	1.141	0.343	15.817
	ALPHA ex COND_TOTAL_FT	0.882	4.767	1.141	0.347	15.868
	ALPHA ex COND_ALPHA	0.833	4.699	1.075	0.338	13.797
	COND_MAX_FT ex D_ALPHA	0.823	4.966	1.057	0.356	13.442
	COND_MAX_FT ex MAX_ALL	0.852	5.090	1.088	0.360	14.345
COND_MAX_FT	COND_MAX_FT ex ALPHA	0.811	4.944	1.045	0.355	13.105
	COND_MAX_FT ex COND_TOTAL_FT	0.841	4.989	1.084	0.361	14.253
	COND_MAX_FT ex COND_ALPHA	0.822	4.924	1.056	0.354	13.379
	COND_TOTAL_FT ex D_ALPHA	0.805	4.915	1.046	0.353	13.106
	COND_TOTAL_FT ex MAX_ALL	0.795	4.941	1.035	0.353	12.811
COND_TOTAL_FT	COND_TOTAL_FT ex ALPHA	0.781	4.856	1.017	0.349	12.323
	COND_TOTAL_FT ex COND_MAX_FT	0.784	4.738	1.030	0.344	12.636
	COND_TOTAL_FT ex COND_ALPHA	0.803	4.810	1.039	0.347	12.885
	COND_ALPHA ex D_ALPHA	0.876	4.900	1.112	0.359	15.068
	COND_ALPHA ex MAX_ALL	0.840	4.687	1.078	0.346	13.936
COND_ALPHA	COND_ALPHA ex ALPHA	0.825	4.791	1.046	0.354	13.113
	COND_ALPHA ex COND_MAX_FT	0.881	4.834	1.126	0.350	15.429
	COND_ALPHA ex COND_TOTAL_FT	0.885	4.873	1.130	0.353	15.580

Table 8: Summary statistics of macro diversified hedge fund data

Table 8 presents summary statistics for the sample of macro diversified hedge funds from 1996 to 2018.

	Mean	**Median**	**10th pct**	**90th pct**	**Std Dev**
Monthly Excess Return (%)	0.114	0.210	-4.247	4.386	4.752
Assets Under Management (AUM, $Millions)	235.770	59.000	4.155	539.665	519.216
Minimum Investment ($Millions)	1.179	0.100	0.028	1.000	4.998
Management Fee (%)	1.480	1.500	0.300	2.000	0.841
Performance Fee (%)	16.956	20.000	0.000	20.000	7.811
Hurdle Rate (%)	0.249	0.000	0.000	0.000	1.380
Lockup Period (days)	32.282	0.000	0.000	0.000	107.815
Redemption Notice (days)	14.046	0.000	0.000	35.000	23.569
Longevity (months)	97.084	78.000	52.000	175.000	52.354

Table 9: In-sample measure values for macro diversified hedge funds

Table 9 presents in-sample statistics for the measures that demonstrate predictive power in the out-of-sample test in the eight-factor timing simulation in Table 3. The mean and standard deviation of each measure are calculated over rolling four-year windows for each macro diversified hedge fund over the sample period, based on the Fung and Hsieh (2004) eight-factor model (15). We sort all in-sample values into quartiles and present the average measure value in each quartile. Q4 represents the average of the highest measure value, and Q1 represents the average of the lowest. The last column presents the *t*-statistic for the difference in means between Q4 and Q1.

Variable	Mean	St Dev	Q4	Q3	Q2	Q1	Q4-Q1 t-stat
D_ALPHA	0.048	0.324	0.436	0.110	-0.023	-0.334	26.173
ALPHA	0.162	0.679	0.959	0.301	-0.006	-0.606	23.881
COND_MAX_FT	0.349	0.514	0.876	0.300	0.144	0.062	10.170
COND_VOL	0.021	0.174	0.211	0.045	-0.042	-0.147	15.090
COND_ALPHA	0.105	0.648	0.738	0.274	-0.056	-0.606	15.130

Table 10: Out-of-sample performance for macro diversified hedge funds

Table 10 presents out-of-sample performance of portfolios based on top and bottom quartiles of past values of each measure. Each portfolio is rebalanced annually based on the updated ranking with respect to preceding four-year window measure values. For unconditional measures, D_ALPHA and ALPHA, portfolios are initially formed on December 31, 1999. For conditional measures, COND_MAX_FT, COND_VOL, and COND_ALPHA, portfolios are initially formed on December 31, 2001, given that calculating the conditional measures requires 24-month history data to calculate past 24-month beta exposures. We invest $1 in each fund within a portfolio at the beginning of the year and track its performance until December 31, 2018. When a fund stops reporting, we assume the fund is defunct and redistribute the remaining capital in the fund equally among surviving portfolio funds. The top-bottom portfolio is equivalent to taking a long position in the top quartile portfolio and a short position in the bottom quartile portfolio. The reported alpha and its corresponding T-value are from the Fung and Hsieh (2004) eight-factor model (15). Return is the mean excess return. The End Value is the dollar value of the initial $1 investment as of December 2018. The Attrition Rate is the average annual rate at which funds stop reporting.

Variable	Portfolio	Alpha	T-Alpha	Return	Sharpe	End Value	Attrition Rate
D_ALPHA	Top	0.272	1.774	0.452	0.167	3.463	4.38%
	Bottom	-0.146	-0.991	-0.027	-0.011	1.186	15.77%
	Top-Bottom	0.417	2.132	0.478	0.166	2.704	-
ALPHA	Top	0.072	0.485	0.241	0.092	2.160	6.43%
	Bottom	0.207	1.457	0.357	0.126	2.776	12.44%
	Top-Bottom	-0.135	-0.813	-0.116	-0.045	0.712	-
COND_MAX_FT	Top	-0.156	-0.733	0.087	0.025	1.297	15.97%
	Bottom	-0.022	-0.198	0.245	0.113	1.928	9.80%
	Top-Bottom	-0.134	-0.619	-0.158	-0.053	0.662	-
COND_VOL	Top	-0.174	-0.944	0.040	0.013	1.209	11.32%
	Bottom	-0.014	-0.082	0.152	0.059	1.565	14.31%
	Top-Bottom	-0.160	-0.667	-0.112	-0.034	0.710	-
COND_ALPHA	Top	0.037	0.316	0.170	0.094	1.678	10.83%
	Bottom	0.298	1.497	0.554	0.154	3.329	15.10%
	Top-Bottom	-0.261	-1.261	-0.385	-0.123	0.412	-

Table 11: Summary statistics of flexible portfolio allocation mutual fund data

Table 11 presents summary statistics for the sample of flexible portfolio allocation mutual funds from 1994 to 2018.

	Mean	Median	10th pct	90th pct	Std Dev
Monthly Net Return (%)	0.427	0.610	-3.255	3.946	3.339
Monthly Gross Return (%)	0.514	0.690	-3.165	4.036	3.340
Total Net Assets (TNA, $Millions)	2091.478	202.000	18.100	3904.807	7505.951
Expense Ratio (%)	1.181	1.150	0.566	1.860	0.525
Turnover Ratio (%)	102.347	58.000	14.000	208.000	180.450
Age (months)	183.329	124.000	40.000	365.500	188.256

Table 12: In-sample measure values for flexible portfolio allocation mutual funds

Table 12 presents in-sample statistics for the measures that demonstrate predictive power in the out-of-sample test in the five-factor timing simulation in Table 6. The mean and standard deviation of each measure are calculated over rolling four-year windows for each flexible portfolio allocation mutual fund over the sample period, based on the five-factor model (16). We sort all in-sample values into quartiles and present the average measure value in each quartile. Q4 represents the average of the highest measure value, and Q1 represents the average of the lowest. The last column presents the *t*-statistic for the difference in means between Q4 and Q1. Panel A presents results based on mutual fund net returns. Panel B presents results based on gross returns. Gross returns are calculated by adding monthly expenses (annual expense ratio divided by 12) to mutual fund net returns.

Panel A: ***Statistics Based on Net Returns***

Variable	Mean	St Dev	Q4	Q3	Q2	Q1	Q4-Q1 t-stat
D_ALPHA	-0.001	0.152	0.160	0.034	-0.023	-0.176	32.339
MAX_ALL	-0.023	0.125	0.122	0.000	-0.050	-0.166	36.614
ALPHA	-0.109	0.310	0.243	-0.025	-0.168	-0.485	39.601
COND_MAX_FT	0.046	0.061	0.116	0.042	0.020	0.005	19.940
COND_TOTAL_FT	-0.003	0.115	0.119	0.013	-0.028	-0.117	24.276
COND_ALPHA	-0.109	0.333	0.197	-0.025	-0.158	-0.459	23.284

Panel B: ***Statistics Based on Gross Returns***

Variable	Mean	St Dev	Q4	Q3	Q2	Q1	Q4-Q1 t-stat
D_ALPHA	-0.001	0.152	0.160	0.034	-0.023	-0.176	32.261
MAX_ALL	-0.024	0.125	0.121	0.000	-0.050	-0.166	36.542
ALPHA	-0.019	0.311	0.334	0.060	-0.078	-0.394	39.327
COND_MAX_FT	0.046	0.061	0.116	0.043	0.020	0.006	20.034
COND_TOTAL_FT	-0.003	0.115	0.119	0.013	-0.028	-0.118	24.270
COND_ALPHA	-0.021	0.332	0.286	0.056	-0.073	-0.361	22.706

Table 13: Out-of-sample performance for flexible portfolio allocation mutual funds

Table 13 presents out-of-sample performance of portfolios based on top and bottom quartiles of past values of each measure. Each portfolio is rebalanced annually based on the updated ranking with respect to preceding four-year window measure values. For unconditional measures, D_ALPHA, MAX_ALL, and ALPHA, portfolios are initially formed on December 31, 1997. For conditional measures, COND_MAX_FT, COND_TOTAL_FT, and COND_ALPHA, portfolios are initially formed on December 31, 1999, given that calculating the conditional measures requires 24-month history data to calculate past 24-month beta exposures. We invest $1 in each fund within a portfolio at the beginning of the year and track its performance until December 31, 2018. When a fund stops reporting, we assume the fund is defunct and redistribute the remaining capital in the fund equally among surviving portfolio funds. The top-bottom portfolio is equivalent to taking a long position in the top quartile portfolio and a short position in the bottom quartile portfolio. The reported alpha and its corresponding T-value are from the five-factor model (16). Return is the mean excess return. The End Value is the dollar value of the initial $1 investment as of December 2018. The Attrition Rate is the average annual rate at which funds stop reporting. Panel A presents results based on mutual fund net returns. Panel B presents results based on gross returns. Gross returns are calculated by adding monthly expenses (annual expense ratio divided by 12) to mutual fund net returns.

Panel A: ***Out-of-Sample Results for Flexible Portfolio Allocation Mutual Funds, Net Returns***

Variable	Portfolio	Alpha	T-Alpha	Return	Sharpe	End Value	Attrition Rate
	Top	-0.005	-0.062	0.301	0.113	2.897	2.31%
D_ALPHA	Bottom	-0.144	-2.359	0.197	0.066	2.180	4.43%
	Top-Bottom	0.140	1.817	0.104	0.082	1.275	-
	Top	-0.181	-2.440	0.150	0.056	1.973	4.26%
MAX_ALL	Bottom	-0.022	-0.287	0.302	0.101	2.834	4.57%
	Top-Bottom	-0.159	-1.865	-0.153	-0.105	0.663	-
	Top	-0.039	-0.599	0.284	0.099	2.735	1.89%
ALPHA	Bottom	-0.160	-1.954	0.132	0.051	1.902	6.83%
	Top-Bottom	0.121	1.342	0.153	0.104	1.430	-
	Top	-0.075	-0.975	0.224	0.082	2.069	4.02%
COND_MAX_FT	Bottom	-0.094	-1.471	0.207	0.079	2.004	2.99%
	Top-Bottom	0.019	0.231	0.017	0.014	1.023	-
	Top	-0.089	-1.183	0.189	0.067	1.898	5.72%
COND_TOTAL_FT	Bottom	-0.016	-0.192	0.269	0.097	2.281	3.53%
	Top-Bottom	-0.073	-0.844	-0.080	-0.063	0.819	-
	Top	0.009	0.133	0.317	0.110	2.528	2.43%
COND_ALPHA	Bottom	-0.148	-1.562	0.143	0.055	1.736	6.16%
	Top-Bottom	0.158	1.389	0.175	0.099	1.436	-

Table 13: Out-of-sample performance for flexible portfolio allocation mutual funds (cont.)

Panel B: ***Out-of-Sample Results for Flexible Portfolio Allocation Mutual Funds, Gross Returns***

Variable	Portfolio	Alpha	T-Alpha	Return	Sharpe	End Value	Attrition Rate
D_ALPHA	Top	0.094	1.269	0.401	0.151	3.717	2.31%
	Bottom	-0.056	-0.914	0.286	0.096	2.725	4.43%
	Top-Bottom	0.151	1.955	0.115	0.090	1.308	-
MAX_ALL	Top	-0.073	-1.019	0.258	0.095	2.586	3.84%
	Bottom	0.075	0.998	0.399	0.134	3.612	4.32%
	Top-Bottom	-0.148	-1.765	-0.141	-0.099	0.682	-
ALPHA	Top	0.065	0.990	0.389	0.136	3.555	2.20%
	Bottom	-0.045	-0.545	0.244	0.096	2.525	6.09%
	Top-Bottom	0.110	1.185	0.145	0.097	1.401	-
COND_MAX_FT	Top	0.033	0.427	0.340	0.121	2.682	4.02%
	Bottom	-0.008	-0.132	0.287	0.111	2.407	2.41%
	Top-Bottom	0.042	0.520	0.054	0.044	1.112	-
COND_TOTAL_FT	Top	-0.012	-0.169	0.267	0.094	2.266	5.28%
	Bottom	0.058	0.666	0.341	0.121	2.681	3.75%
	Top-Bottom	-0.070	-0.803	-0.073	-0.057	0.830	-
COND_ALPHA	Top	0.060	0.842	0.362	0.123	2.788	2.43%
	Bottom	-0.041	-0.419	0.256	0.097	2.240	6.25%
	Top-Bottom	0.101	0.872	0.106	0.058	1.225	-

Table 14: Correlations among D_ALPHA, ALPHA, and COND_ALPHA

Table 14 presents correlations among D_ALPHA, ALPHA, and COND_ALPHA. Panel A presents results for hedge funds; Panel B presents results for mutual funds. All measures are calculated over rolling four-year windows over the sample period, based on the Fung and Hsieh (2004) eight-factor model (15), for hedge funds in Panel A, and on the five-factor model (16), for mutual funds in Panel B. P-values are presented below the correlation coefficients.

Panel A: ***Correlations in Hedge Funds***

	D_ALPHA × ALPHA	D_ALPHA × COND_ALPHA	ALPHA × COND_ALPHA
Simulated Timers	0.757	0.525	0.754
	<0.001	<0.001	<0.001
Macro Diversified Hedge Funds	0.120	0.021	0.887
	0.006	0.705	<0.001

Panel B: ***Correlations in Mutual Funds***

	D_ALPHA × ALPHA	D_ALPHA × COND_ALPHA	ALPHA × COND_ALPHA
Simulated Timers	0.792	0.583	0.865
	<0.001	<0.001	<0.001
Flexible Allocation Mutual Funds	0.363	0.348	0.935
	<0.001	<0.001	<0.001

Appendix

Table A1: Summary statistics from the eight-factor timing simulation (cont.)

Table A1 presents standard deviations of proposed timing measure values for the nine types of imperfect timers. All measures are based on the Fung and Hsieh (2004) eight-factor model (15). The imperfect timers make errors in both timing of switching into the optimal factor and in the optimal factor selection. We simulate three levels of accuracy in both dimensions, which yields nine possible combinations in timing accuracy combined with factor selection accuracy: we simulate 10,000 timers for each combination. We simulate timing accuracy by allowing the factor switch time within two, four, and six months prior to the optimal switch time (the optimal switch time included) via uniform distribution. We simulate factor selection accuracy in that the optimal factor at each switch point is selected with probabilities of 90%, 60%, and 30%. For each timer, we calculate each measure over four-year rolling windows over the sample period of 25 years from January 1994 to December 2018. This yields 21,000 (21*10,000 timers) observations for each category of accuracy for the unconditional measures, D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, and 19,000 (19*10,000 timers) observations for conditional measures, COND_MAX_FT, COND_TOTAL_FT, COND_VOL, COND_ALPHA, given that calculating the conditional measures requires 24-month history data to calculate past 24-month beta exposures.

Timing Accuracy		**Factor Selection Accuracy**		
		90%	**60%**	**30%**
2 months	D_ALPHA	0.904	0.938	0.904
	MAX_ALL	0.239	0.250	0.237
	MAX_SINGLE	2.405	2.169	1.940
	SUM_ALL	4.542	4.165	3.779
	ALPHA	1.547	1.660	1.593
	COND_MAX_FT	1.432	5.336	6.958
	COND_TOTAL_FT	1.125	1.641	3.209
	COND_VOL	0.584	1.367	3.051
	COND_ALPHA	1.563	1.501	1.402
4 months	D_ALPHA	0.819	0.869	0.850
	MAX_ALL	0.240	0.245	0.230
	MAX_SINGLE	2.345	2.108	1.901
	SUM_ALL	4.534	4.118	3.721
	ALPHA	1.399	1.497	1.453
	COND_MAX_FT	1.201	4.047	82.791
	COND_TOTAL_FT	1.143	1.240	11.034
	COND_VOL	0.762	0.949	10.781
	COND_ALPHA	1.415	1.363	1.289
6 months	D_ALPHA	0.782	0.836	0.820
	MAX_ALL	0.236	0.236	0.221
	MAX_SINGLE	2.250	2.028	1.852
	SUM_ALL	4.603	4.160	3.749
	ALPHA	1.457	1.464	1.395
	COND_MAX_FT	73.646	4.820	133.635
	COND_TOTAL_FT	10.828	1.158	18.853
	COND_VOL	10.571	0.856	18.444
	COND_ALPHA	1.373	1.294	1.223

Table A2: Summary statistics from the five-factor timing simulation (cont.)

Table A2 presents standard deviations of proposed timing measure values for the nine types of imperfect timers. All measures are based on the five-factor model (16). The imperfect timers make errors in both timing of switching into the optimal factor and in the optimal factor selection. We simulate three levels of accuracy in both dimensions, which yields nine possible combinations in timing accuracy combined with factor selection accuracy: we simulate 10,000 timers for each combination. We simulate timing accuracy by allowing the factor switch time within two, four, and six months prior to the optimal switch time (the optimal switch time included) via uniform distribution. We simulate factor selection accuracy in that the optimal factor at each switch point is selected with probabilities of 90%, 60%, and 30%. For each timer, we calculate each measure over four-year rolling windows over the sample period of 25 years from January 1994 to December 2018. This yields 21,000 (21*10,000 timers) observations for each category of accuracy for the unconditional measures, D_ALPHA, MAX_ALL, MAX_SINGLE, SUM_ALL, ALPHA, and 19,000 (19*10,000 timers) observations for conditional measures, COND_MAX_FT, COND_TOTAL_FT, COND_VOL, and COND_ALPHA, given that calculating the conditional measures requires 24-month history data to calculate past 24-month beta exposures.

Timing Accuracy		**Factor Selection Accuracy**		
		90%	**60%**	**30%**
2 months	D_ALPHA	0.352	0.335	0.302
	MAX_ALL	0.439	0.385	0.325
	MAX_SINGLE	0.482	0.512	0.489
	SUM_ALL	1.453	1.335	1.194
	ALPHA	0.707	0.659	0.572
	COND_MAX_FT	0.313	0.288	0.231
	COND_TOTAL_FT	0.399	0.366	0.326
	COND_VOL	0.220	0.223	0.215
	COND_ALPHA	0.568	0.562	0.504
4 months	D_ALPHA	0.334	0.322	0.292
	MAX_ALL	0.449	0.388	0.322
	MAX_SINGLE	0.470	0.506	0.494
	SUM_ALL	1.536	1.391	1.230
	ALPHA	0.665	0.610	0.525
	COND_MAX_FT	0.325	0.326	0.302
	COND_TOTAL_FT	0.417	0.404	0.343
	COND_VOL	0.235	0.258	0.229
	COND_ALPHA	0.554	0.519	0.457
6 months	D_ALPHA	0.324	0.311	0.279
	MAX_ALL	0.420	0.364	0.302
	MAX_SINGLE	0.468	0.496	0.483
	SUM_ALL	1.465	1.332	1.180
	ALPHA	0.609	0.552	0.472
	COND_MAX_FT	0.261	0.300	0.197
	COND_TOTAL_FT	0.380	0.380	0.308
	COND_VOL	0.215	0.246	0.191
	COND_ALPHA	0.542	0.482	0.412