Measuring the Effects of Foresight and Commitment on Portfolio Performance

by

Kenneth Khang
College of Business
Idaho State University
Pocatello, ID 83209
khankenn@isu.edu

and

Thomas W. Miller, Jr.[1]
John Cook School of Business
Saint Louis University
3674 Lindell Boulevard
St. Louis, MO 63108
PH: (314) 977-3830
millertw@slu.edu

First draft: December 16, 2007

Current draft: August 16, 2009

[1] Contact author. We thank Leigh Riddick for providing weight allocation data and acknowledge Mary Bange for her contributions on related collaborations. Mike Highfield, Carolyn Moore, and seminar participants at Mississippi State University provided helpful comments and suggestions.

Measuring the Effects of Foresight and Commitment on Portfolio Performance

Abstract

A central question in finance is how to measure portfolio performance. We assert that foresight (i.e., the ability of managers to forecast relative returns) and commitment (i.e., the aggressiveness with which managers act on their foresight) are necessary ingredients to generate portfolio performance not attributable to chance. We create two portfolio performance measures free from the need to pre-specify a benchmark portfolio. Our measures allow for many investment styles because they use benchmarks created from the style implied in the portfolio being tested. To calibrate these new measures, we use a weight-generating process that simulates various combinations of foresight and commitment. We decompose our measures to show how foresight and commitment act jointly to generate statistically measurable excess performance. An outcome of our measures is a scoring mechanism that groups portfolios by commitment, and then ranks portfolio performance by foresight. We apply our measures to a series of actual asset allocation recommendations made by a panel of Investment Houses. Largely, we find little foresight, but we document considerable aggression. The Investment Houses made allocation recommendations to equity, bonds, and cash. However, our measures easily generalize to a wide variety of potentially hard to measure investment styles, like those of hedge funds and other alternative investment managers.

Measuring the Effects of Foresight and Commitment on Portfolio Performance

1. Introduction

Perhaps the most fundamental investment decision is how best to allocate portfolio dollars among alternatives. Today, investment managers have a wide variety of equity, fixed income, real estate, precious metal, and commodity index funds from which to choose. The focus of our investigation is on how to measure portfolio performance when an investment manager actively changes his asset allocation among equity, bonds, and cash. However, the methods we introduce easily generalize to a wide variety of potentially hard to measure investment styles, like those of hedge funds and other alternative investment managers.[2]

Beginning with Jensen (1968), most research on portfolio performance measurement focuses on measuring the performance of equity portfolios. In addition, much of this research implicitly assumes that managers maintain a constant portfolio risk level. However, as Jensen (1968) writes:

> *"…the portfolio manager can certainly change the risk level of his portfolio very easily [by] consciously switch[ing] portfolio holdings between equities, bonds, and cash in trying to outguess the movement of the market."*

Similarly, most of the research on asset allocation performance uses a constant blend of each asset class [Arshanapalli, Coggin, and Nelson (2001) and Brinson et. al. (1986, 1991).] Clearly, these restrictive approaches ignore variation in the investment style and risk of the portfolio due to shifts in asset allocation percentages.

We create two portfolio performance measures that account for investment style and risk. The first measure is a modified version of the randomized benchmark introduced by

[2]Astute asset allocation among equity, bonds, and cash can generate substantial profits. For example, from January 1980 through December 2008, an investment of $1 million would have grown to over $28.4 billion if, at the beginning of each month, the investor placed their growing portfolio into the asset class (i.e., equities, bonds, or cash) that will perform best over the month (with a transactions costs of 50 basis points per month, the portfolio still would have grown to about $5 billion). By contrast, an investment of $1 million in the S&P 500 would have grown to $19.2 million over this time period (assuming total returns, i.e., dividends included). Blake, Lehmann, and Timmermann (1999) underscore the importance of asset allocation when they show that potential gains to investors from asset allocation dwarf potential gains from market-timing and security selection.

Bange, Khang, and Miller (2008). The second measure is a modified version of the unconditional weight-based measure of Ferson and Khang (2002) and Grinblatt and Titman (1993).

These two portfolio performance measures: 1) do not use pre-specified benchmark portfolios; and, 2) do not assume the manager keeps a constant portfolio risk level. Our measures have the desirable quality of sharing the style and risk profiles of the measured portfolio throughout the portfolio's holding period. In our investigation, we focus on portfolio weight allocations across indexes for equity, bonds, and cash. However, it is possible to generalize our technique to many asset classes. The only inputs needed for our measures are beginning-of-period portfolio weights and subsequent asset-class returns.

Our method creates, versus pre-specifies, a benchmark portfolio that shares a similar style and risk profile with the measured portfolio. We define outperformance as portfolio returns in excess of the benchmark portfolio, and our measures capture this difference. Further, we decompose outperformance into two separate, intuitive components and extend our measures to capture them. The two components we investigate are: 1) Foresight, i.e., how well can managers forecast relative returns given their information set? and, 2) Commitment, i.e., how aggressively do managers act on their foresight?

Recent research buttresses our assertion that foresight and commitment are necessary aspects of measureable outperformance. Kacperczyk, Sialm, and Zheng (2005) find evidence that equity mutual fund managers exhibiting industry concentration in their portfolios outperform common equity benchmarks. Ivkovich, Sialm, and Weisbenner (2008) find that individual investors who hold concentrated portfolios outperform investors holding more diversified portfolios. These results suggest that those with foresight in selecting stocks and managing an equity portfolio tend to exhibit commitment by holding less diversified portfolios.

However, the research questions we explore here focus on active asset allocation instead of equity selection. Our first question is: How do we measure the performance of a portfolio with an active asset allocation in the face of changing style and risk? Our second question is: How do we measure foresight and commitment separately? That is, 1) how closely

do the portfolio weights move with future returns (foresight); and, 2) how large are the changes in portfolio weights (commitment) that take advantage of foresight?

Using simulated portfolio weight sets endowed with varied combinations of foresight and commitment, we find that both performance measures have power to detect performance. Further, the randomized and the weight-based measures generate consistent results.

We show that our foresight measure is well-behaved because this measure captures a manager's ability to allocate weights among assets *ex ante*. That is, our foresight measure estimates the contribution to outperformance resulting from the tendency of the manager to increase (decrease) the weight on an asset above the benchmark weight when future returns are higher (lower) relative to the other assets.

We find that our commitment measure captures the magnitude of weight differences from the benchmark with enough precision to allow us to create a scoring mechanism that groups portfolios by commitment.[3] Our scoring mechanism has important implications for investors who rely on basic asset allocation advice. Conceivably, active investment managers can be scored along foresight and commitment dimensions.[4]

As an application of our randomized and our weight-based measures, we examine the performance of actual Investment House recommendations presented in Bange, Khang, and Miller (2008). We examine performance and score the Investment Houses for foresight and commitment. We find that the Houses fail to generate positive outperformance using our randomized and weight-based measures. No House generated significant foresight. Their commitment measures, however, are generally higher than the commitment levels of the simulated portfolios—suggesting that the Houses strove for performance through aggressive weight changes.

The paper proceeds as follows. Section 2 contains a brief survey of related research, the motivation for our performance measures, and the description of our measures of foresight

[3] Note that commitment will be a function of risk aversion and foresight. Risk aversion should affect a manager's propensity to make changes to the portfolio in response to a given information signal. Of course, the nature of the change will depend on the utility function. Foresight may affect commitment as well. For example, if the forecasted return to equity is much higher than to the other asset classes, then the shift to equity may be larger than if the return advantage forecasted for equity were relatively smaller.

[4] Mutual fund investments in particular are vast. As of year-end 2008, worldwide mutual fund assets totaled $18.97 trillion, distributed among 69,032 funds across 45 countries (www.ici.org/pdf/2009_factbook.pdf).

and commitment. We describe our weight-generating process and discuss the characteristics of the simulated portfolios it creates in Section 3. In Section 4, we apply our measures of performance, foresight, and commitment to the simulated portfolios to illustrate the behavior of the measures. In Section 5, we re-examine the performance of the Investment Houses examined by Bange, Khang, and Miller (2008) using our measures. We conclude with Section 6.

2. Methods

2.1 Related Research

A central question in finance is how to measure portfolio performance. The academic literature on this question has two primary approaches: returns-based measurement and weights-based portfolio performance measurement.

Returns-based portfolio measurement generally judges a portfolio by comparing its returns to the returns of a risk-adjusted benchmark portfolio. Jensen (1968) led early inquiries into returns-based portfolio performance. Using the CAPM of Sharpe (1964), Jensen (1968) creates the concept of Alpha--now a long-standing measure of a portfolio's return in excess of its expected return. Although based on an asset-pricing model, the Jensen (1968) technique is subject to the issue that arises if one uses an inefficient benchmark portfolio to rank performance [Roll (1978)]. Briefly, Roll (1978) shows that it is possible to reorder portfolio performance rankings generated from a mean-variance inefficient benchmark portfolio by using another mean-variance inefficient portfolio.[5]

Using an "Alpha" method to measure portfolio performance has broad appeal—despite Roll's (1978) critiques. Over the years, this method has evolved from using one factor. Fama and French (1993) introduce an empirically based three-factor model that Carhart (1997) extends to four factors. Even though using these three- and four-factor models has become widespread, they share the nettlesome feature that they are not based upon an asset-pricing model.

In the realm of portfolio performance measurement, users of multi-factor models generally assume that a manager's style and the level of risk taken by the manager are constant

[5]Alas, Roll (1977) shows that one cannot simply use an efficient mean-variance portfolio because the mean-variance efficient portfolio is unobservable.

over the holding period. However, Ferson and Schadt (1996) show that conclusions about a portfolio manager's skill can change after making adjustments for style shifts and risk changes. Wermers (2000) reports style differences between the top-quintile and the bottom-quintile mutual funds in his extensive study. Recently, Bollen and Whaley (2009) allow the portfolio's risk-level to shift once during the holding period. Bollen and Whaley (2009) find that about half the hedge funds in their sample exhibit statistically significant shifts in their risk exposures.

Weights-based portfolio performance measures begin with the simple concept that skilled portfolio managers will shift their portfolio weights in the same direction as subsequent returns.[6] Copeland and Mayers (1982) and Grinblatt and Titman (1993) develop portfolio performance measures that do not use pre-specified benchmark portfolios. Conceptually, the unconditional weight methods equate a portfolio manager's performance with the unconditional covariance between the portfolio weights and subsequent returns.

Ferson and Khang (2002) create a portfolio performance measure that disentangles public information from a manager's private information. Ferson and Khang (2002) show that it is possible for a manager to exhibit positive unconditional performance while having no conditional performance. They show that this can occur if a portfolio manager uses publicly available information when purchasing or selling securities.

The Ferson and Khang (2002) inference is consistent with a long-standing issue in judging superior performance, which Wermers (2006) summarizes well:

> *"An ideal property of a performance measure is to rank managers by the precision of their private information, [i.e.] skills that cannot be captured through simple strategies…implemented by an uninformed observer."*

To illustrate the issue, suppose a manager's task is to track, but also beat, a pre-specified benchmark portfolio. Roll (1992) shows that the manager can accomplish this task by selecting a portfolio with a beta greater than one—when the beta is estimated using the benchmark. Daniel and Titman (1997) show that a mechanical strategy exists to earn excess returns over a benchmark based on the four-factor Carhart (1997) model that controls for multiple "risks."

[6] Wermers (2006) provides a detailed survey of weight-based portfolio performance.

Daniel and Titman (1997) argue that a strategy of selecting securities that have low covariance with the four factors results in a (false) positive alpha.

The motivation underpinning our measures is the desire to create a method that ranks managers by their total skill—proprietary and mechanical. If a manager fails to outperform via our measures, then the manager's performance is consistent with the performance of a manager who is not employing any mechanical or proprietary strategies. If our measures judge a manager's performance as significant, one could perform further tests, like those of Barras, Scaillet, and Wermers (2009) or Ferson and Khang (2002), to examine other aspects of the nature of the excess portfolio returns.[7]

2.2 The Measures

2.2.1 Measuring Performance

Barras, Scaillet, and Wermers (2009) develop a technique to separate mutual fund performance into luck and skill.[8] However, because Barras, Scaillet, and Wermers (2009) rely on pre-specified factors—their technique is subject to the set of theoretical problems outlined in Wermers (2006). In addition, their technique does not allow them to capture style and risk on an individual fund basis. In fact, their technique assumes constant style and risk-level.[9]

By contrast, our portfolio performance methods capture the investment style and the ex ante risk of the measured portfolio. Moreover, our methods do not use a pre-specifed

[7]Conditional versions of both our measures could directly address the issue summarized in Wermers' quote. Nonetheless, our randomized benchmark design will perform well in many instances, though not all. For example, consider an instance where periodic equity returns are always higher than periodic bond returns along with a manager who increases his equity weights through time. In this case, the manager will appear to have information because all equity weight changes—no matter how we shuffle them—will be positive. The way to correct for this unlikely scenario is to use conditioning information that accounts for the fact that expected equity returns are greater than the expected returns for the other asset classes in the measurement interval.

[8] Barras, Scaillet, and Wermers (2009) generate alphas using Carhart's (1997) four-factor model, i.e., adding momentum (MOM) to the Fama-French factors: overall market return, market cap (SMB), and book-to-market (HML). They also use a conditional four-factor model to account for time-varying exposure to the market portfolio via Ferson and Schadt (1996).

[9] However, it should be noted that Barras, Scaillet, and Wermers (2009) dataset categorizes funds into three pre-specified styles: Growth, Aggressive Growth, and Growth & Income.

benchmark portfolio. Our approach to dealing with style is similar to Ferson and Khang (2002) and Grinblatt and Titman (1993). That is, we also adopt the "style is, as style does" approach.

By using the measured portfolio to create its own benchmark, the benchmark portfolio has a style similar to the measured portfolio along four dimensions. Our benchmark portfolios: 1) are composed of the same assets as the measured portfolio; 2) are subject to weight changes of similar magnitude; 3) contain similar trends in the weights over the measurement period, and; 4) have a similar average exposure to the different asset classes over the measurement period. In addition, by matching investment style in this manner, we conjecture that these benchmark portfolios have approximately the same ex ante risk as the measured portfolio.

We use two methods to create our performance measures. In the first measure, we use a modified version of the randomized shuffled-weight change benchmark portfolio introduced by Bange, Khang, and Miller (2008). Our performance measure is simply the return difference between the measured portfolio and the benchmark portfolio. In the second measure, we modify the unconditional weight-based measure of Ferson and Khang (2002) and Grinblatt and Titman (1993). The main modification we make to this measure is that we demean returns *cross-sectionally* when we calculate the covariance between weight changes and subsequent returns.

2.2.2 Portfolio Performance Defined and Benchmark Weights Created

To motivate our measures and lay the groundwork for defining foresight and commitment, consider the following portfolio performance definition over period t:

$$perf_{p,t} \equiv \tilde{r}_{p,t} - \tilde{r}_{b,t} \qquad (1).$$

In Equation (1), $\widetilde{perf}_{p,t}$ is the performance of the portfolio in period t; $\tilde{r}_{p,t}$ is the period t return of the portfolio; and $\tilde{r}_{b,t}$ is the period t return of the benchmark portfolio. Thus, in this definition, performance is the amount by which the original portfolio outperforms the benchmark in any given period t. Our performance measure is the average, or expected, outperformance. Therefore, we can write

$$E\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} = E\{\sum_{j=1}^{N} \widetilde{w}_{j,t}\, \tilde{r}_{j,t} - \sum_{j=1}^{N} \widetilde{w}_{j,t}^{b}\, \tilde{r}_{j,t}\} \qquad (2).$$

In Equation (2), $\widetilde{w}_{j,t}$ is the weight on asset j in the original portfolio at the beginning of period t; $\tilde{r}_{j,t}$ is the period t return on asset j; and $w_{j,t}^{b}$ is the weight on asset j in the benchmark portfolio at the beginning of period t. Consider the returns to a portfolio of equity, bonds, and cash over a measurement interval of T months. Letting j = e, b, or c, for equity, bonds, and cash, we create matrix M as

$$\mathbf{M} = \begin{bmatrix} w_{e,1}r_{e,1} & w_{b,1}r_{b,1} & w_{c,1}r_{c,1} \\ w_{e,2}r_{e,2} & w_{b,2}r_{b,2} & w_{c,2}r_{c,2} \\ w_{e,3}r_{e,3} & w_{b,3}r_{b,3} & w_{c,3}r_{c,3} \\ \vdots & \vdots & \vdots \\ w_{e,T}r_{e,T} & w_{b,T}r_{b,T} & w_{c,T}r_{c,T} \end{bmatrix} = \begin{bmatrix} w_{e,1} & w_{b,1} & w_{c,1} \\ w_{e,2} & w_{b,2} & w_{c,2} \\ w_{e,3} & w_{b,3} & w_{c,3} \\ \vdots & \vdots & \vdots \\ w_{e,T} & w_{b,T} & w_{c,T} \end{bmatrix} \# \begin{bmatrix} r_{e,1} & r_{b,1} & r_{c,1} \\ r_{e,2} & r_{b,2} & r_{c,2} \\ r_{e,3} & r_{b,3} & r_{c,3} \\ \vdots & \vdots & \vdots \\ r_{e,T} & r_{b,T} & r_{c,T} \end{bmatrix} \qquad (3).$$

In Equation (3), # represents element-wise multiplication; $w_{j,t}$ is the weight on asset j at the beginning of period t; and $r_{j,t}$ is the return to asset j over period t. We use periodic portfolio returns, i.e., the sums across the rows of matrix M, to calculate annualized geometric returns.

2.2.3 The Randomized Lagged-Weight Measure (RLM)

Our first measure is the randomized lagged-weight measure (RLM). To create the RLM, we use the randomization concept from Bange, Khang, and Miller (2008) to create the benchmark in Equation (2). However, we combine the randomization with the idea that using lagged weights keeps the benchmark style close to the style of the original portfolio [see Ferson and Khang (2002) and Grinblatt and Titman (1993)][10].

We create the randomized benchmark, or $\tilde{r}_{b,t}$, in Equation (1), for the randomized lagged-weight measure (RLM) as follows. First, we assume an endowment weight set at time 0. Then, we calculate weight changes as follows.

[10] The Bange, Khang, and Miller (2008) benchmark randomly adds an actual change to the weight already chosen for the upcoming period. They show that their benchmark maintains risk characteristics similar to the original portfolio.

$$\mathbf{D} = \begin{bmatrix} d_{e,1} & d_{b,1} & d_{c,1} \\ d_{e,2} & d_{b,2} & d_{c,2} \\ d_{e,3} & d_{b,3} & d_{c,3} \\ \vdots & \vdots & \vdots \\ d_{e,T} & d_{b,T} & d_{c,T} \end{bmatrix} = \begin{bmatrix} w_{e,1} & w_{b,1} & w_{c,1} \\ w_{e,2} & w_{b,2} & w_{c,2} \\ w_{e,3} & w_{b,3} & w_{c,3} \\ \vdots & \vdots & \vdots \\ w_{e,T} & w_{b,T} & w_{c,T} \end{bmatrix} - \begin{bmatrix} w_{e,0} & w_{b,0} & w_{c,0} \\ w_{e,1} & w_{b,1} & w_{c,1} \\ w_{e,2} & w_{b,2} & w_{c,2} \\ \vdots & \vdots & \vdots \\ w_{e,T} & w_{b,T} & w_{c,T} \end{bmatrix} \quad (4),$$

where $d_{j,t}$ is the weight change on asset j from the beginning of period t to the beginning of period t+1.

The randomization focuses on the question of whether the order of portfolio changes made by the manager matters. The randomized benchmark weights, used for $\tilde{r}_{b,t}$ in Equation (1), are created by substituting the changes that a manager actually made to his portfolio with randomly selected actual changes the manager made at other times during the measurement period.[11]

To create the randomized benchmark for the RLM, we shuffle the rows of matrix D and then add them to the lagged weight matrix. As an example, suppose that after the shuffle, the $d_{j,14}$'s are now in row 2, $d_{j,9}$'s are now in row 3, and $d_{j,7}$'s are now in row T. Thus,

$$W^b = \begin{bmatrix} w^b_{e,1} & w^b_{b,1} & w^b_{c,1} \\ w^b_{e,2} & w^b_{b,2} & w^b_{c,2} \\ w^b_{e,3} & w^b_{b,3} & w^b_{c,3} \\ \vdots & \vdots & \vdots \\ w^b_{e,T} & w^b_{b,T} & w^b_{c,T} \end{bmatrix} = \begin{bmatrix} w_{e,0} & w_{b,0} & w_{c,0} \\ w_{e,1} & w_{b,2} & w_{c,2} \\ w_{e,2} & w_{b,3} & w_{c,3} \\ \vdots & \vdots & \vdots \\ w_{e,T-1} & w_{b,T-1} & w_{c,T-1} \end{bmatrix} + \begin{bmatrix} d_{e,5} & d_{b,5} & d_{c,5} \\ d_{e,14} & d_{b,14} & d_{c,14} \\ d_{e,9} & d_{b,9} & d_{c,9} \\ \vdots & \vdots & \vdots \\ d_{e,7} & d_{b,7} & d_{c,7} \end{bmatrix} \quad (5).$$

and

$$\mathbf{M^b} = \begin{bmatrix} W^b_{e,1} r_{e,1} & W^b_{b,1} r_{b,1} & W^b_{c,1} r_{c,1} \\ W^b_{e,2} r_{e,2} & W^b_{b,2} r_{b,2} & W^b_{c,2} r_{c,2} \\ W^b_{e,3} r_{e,3} & W^b_{b,3} r_{b,3} & W^b_{c,3} r_{c,3} \\ \vdots & \vdots & \vdots \\ W^b_{e,T} r_{e,T} & W^b_{b,T} r_{b,T} & W^b_{c,T} r_{c,T} \end{bmatrix} = \begin{bmatrix} w^b_{e,1} & w^b_{b,1} & w^b_{c,1} \\ w^b_{e,2} & w^b_{b,2} & w^b_{c,2} \\ w^b_{e,3} & w^b_{b,3} & w^b_{c,3} \\ \vdots & \vdots & \vdots \\ w^b_{e,T} & w^b_{b,T} & w^b_{c,T} \end{bmatrix} \# \begin{bmatrix} r_{e,1} & r_{b,1} & r_{c,1} \\ r_{e,2} & r_{b,2} & r_{c,2} \\ r_{e,3} & r_{b,3} & r_{c,3} \\ \vdots & \vdots & \vdots \\ r_{e,T} & r_{b,T} & r_{c,T} \end{bmatrix} \quad (6).$$

[11]Note that the rows in W^b may not sum to one in cases where the change forces a negative weight. In those cases we proportionately rebalance the rows in W^b to ensure the benchmark weights sum to one.

As we do with Matrix M, we use the sums across the rows of matrix M^b, to calculate annualized geometric returns for the randomized benchmark. In our empirical tests, we repeat the process of shuffling the weights and calculating returns 10,000 times. Thus, the randomized lagged-weight measure (RLM) is

$$\text{RLM} = \text{E}\{\text{actual portfolio return} - \text{randomized benchmark return}\}. \qquad (7).$$

The weights used for the randomized benchmark in the RLM are similar to the original portfolio in four ways. First, the randomized benchmark is composed of the same assets as the original portfolio. Second, the randomized benchmark is subject to weight changes of similar magnitude, because it shuffles the weight changes the manager actually made over the measurement period. Third, the randomized benchmark contains similar trends in the weights over the measurement period because it uses the lagged weights. For example, if the equity weight rises from 20% at the beginning of the measurement period to 80% by the end, then a similar pattern will be preserved in the randomized benchmark. Fourth, the randomized benchmark has similar average exposure to the different asset classes over the measurement period as the original portfolio.

2.2.4 The Weight-Based Cross-Sectional Return Measure (WCM)

Our second measure uses a cross-sectional return version of the unconditional weight-based measure from Ferson and Khang (2002) and Grinblatt and Titman (1993). This measure uses the lagged weights of the original portfolio as the benchmark weights, though this is not necessarily the only choice.[12] Conceptually, this weight-based measure estimates the covariance between weight changes (or weight differences from a benchmark) and subsequent returns. We call this measure the weight-based cross-sectional return measure (WCM). The WCM is calculated as

$$WCM = E\left\{\sum_{j=1}^{N}\left(\widetilde{w}_{j,t} - \widetilde{w}_{j,t-1}\right)\left(\tilde{r}_{j,t} - \bar{r}\right)\right\} \qquad (8).$$

The WCM differs from the Unconditional Weight Measure of Ferson and Khang (2002) in two ways. First, similar to Grinblatt and Titman (1993), we use the lagged weight rather than

[12] Ferson and Khang (2002) note that any benchmark weight may be used, not just the lagged portfolio weights.

the buy-and-hold weight as the benchmark. In measuring the performance of an actual portfolio, the buy-and-hold weight is preferable. However, because we apply this measure to a set of actual weight recommendations from a group of Investment Houses, we choose to use lagged weights.

Second, we calculate demeaned returns relative to the cross-sectional mean return. In contrast, Ferson and Khang (2002) use the time series mean for each asset. In our study, we examine active asset allocation strategies where a manager chooses to allocate dollars among a set of asset classes. The cross-sectional mean return is appropriate because the manager's goal at any given period t is to allocate more dollars towards asset classes whose returns are forecasted to be relatively higher than the returns of the other asset classes in the portfolio.[13]

By using demeaned returns relative to the cross-sectional mean return, we can rewrite Equation (8) as

$$\mathrm{WCM} = \mathrm{E}\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} \qquad (9).$$

Equation (9) shows that the WCM is also the average amount by which the original portfolio return exceeds the lagged weight benchmark return.

Lastly, like the RLM, the WCM does not pre-specify an investment style. The benchmark is simply the lagged portfolio. Thus, the benchmark for WCM will be similar to the original portfolio in the same four ways as the benchmark for the RLM is similar to the original portfolio. This ensures that the WCM has similar style and risk characteristics as the original portfolio.

2.3. Decomposing Performance into Foresight, Commitment, and Opportunity

In decomposing portfolio performance into Foresight, Commitment, and Opportunity, we do not make any assumptions about utility functions or information sets. The strength of our decomposition is that we are not assuming any particular utility maximization specification. The weakness of this decomposition is that we cannot make specific statements about how risk tolerance and information set influence foresight and commitment.

[13]In examining a portfolio whose goal is to select among individual securities (in other words, security selection), the manager may be thought of as allocating more dollars toward those securities whose return is forecasted to be higher than required for that particular security. Grinblatt and Titman (1993) do not use demeaned returns, so their measure is not a true covariance [see Ferson and Khang (2002) for a discussion].

To begin, we recall our performance measure in Equation (2):

$$E\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} = E\{\sum_{j=1}^{N} \tilde{w}_{j,t}\, \tilde{r}_{j,t} - \sum_{j=1}^{N} \tilde{w}_{j,t}^{b}\, \tilde{r}_{j,t}\} \qquad (10).$$

Because the sum of the portfolio weights must equal one, we can write

$$E\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} = E\{\sum_{j=1}^{N} \tilde{w}_{j,t}\, \tilde{r}_{j,t} - \sum_{j=1}^{N} \tilde{w}_{j,t}^{b}\, \tilde{r}_{j,t} - \sum_{j=1}^{N} \tilde{w}_{j,t}\, \bar{r} + \sum_{j=1}^{N} \tilde{w}_{j,t}^{b}\, \bar{r}\} \qquad (10a),$$

where $\bar{r}$ is the cross-sectional mean return across assets. Rearranging,

$$E\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} = E\{\sum_{j=1}^{N} (\tilde{w}_{j,t} - \tilde{w}_{j,t}^{b})\, \tilde{r}_{j,t} - \sum_{j=1}^{N} (\tilde{w}_{j,t} - \tilde{w}_{j,t}^{b})\, \bar{r}\} \qquad (10b),$$

$$E\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} = E\{\sum_{j=1}^{N} (\tilde{w}_{j,t} - \tilde{w}_{j,t}^{b})\, (\tilde{r}_{j,t} - \bar{r})\} \qquad (10c).$$

Noting that the cross-sectional mean weight change $\overline{(\tilde{w}_{J,t} - \tilde{w}_{J,t}^{b})} = 0$, we can write

$$E\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} = E\left\{\sum_{j=1}^{N} \left[(\tilde{w}_{j,t} - \tilde{w}_{j,t}^{b}) - \overline{(\tilde{w}_{J,t} - \tilde{w}_{J,t}^{b})}\right] (\tilde{r}_{j,t} - \bar{r})\right\} \qquad (10d).$$

Equation (9d) is the definition of the cross-sectional covariance, i.e.,

$$E\left\{\sum_{j=1}^{N} \left[(\tilde{w}_{j,t} - \tilde{w}_{j,t}^{b}) - \overline{(\tilde{w}_{J,t} - \tilde{w}_{J,t}^{b})}\right] (\tilde{r}_{j,t} - \bar{r})\right\} \equiv Cov(\tilde{w}_{j,t} - \tilde{w}_{j,t}^{b}, r_t) \qquad (10e).$$

Therefore,

$$E\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} = Cov(\tilde{w}_{j,t} - \tilde{w}_{j,t}^{b}, r_t) \qquad (10f),$$

or,

$$E\{\tilde{r}_{p,t} - \tilde{r}_{b,t}\} = \rho_{\Delta w,r,t}\, \sigma(\Delta w_t)\, \sigma(r_t) \qquad (11).$$

In Equation (11), $\rho_{\Delta w,r,t}$ is the cross-sectional correlation between the weight differences and subsequent returns at time t; $\Delta w_t = (\tilde{w}_{j,t} - \tilde{w}_{j,t}^{b})$; $\sigma(\Delta w_t)$ is the standard

deviation of the weight differences across assets at time t; and $\sigma(r_t)$ is the cross-sectional standard deviation of the returns at time t.

Equation (11) contains the terms we use for foresight and commitment. We define $\rho_{\Delta w,r,t}$ as foresight and $\sigma(\Delta w_t)$ as commitment. The third term, $\sigma(r_t)$, can be thought of as opportunity.[14] Conceptually, foresight is the component of outperformance that is due to "predictive" ability. Commitment is the aggressiveness with which one implements foresight and is the component of outperformance that magnifies the effects of any foresight. In fact, we can consider it as a multiplier for foresight. The opportunity term captures the extent to which the patterns of returns were favorable for generating outperformance. Opportunity is also a multiplier of foresight. Opportunity is the component of outperformance attributable to differences in returns among asset classes. As we measure it, opportunity is exogenous, and constant, across portfolios over any given holding period.

We abbreviate our measures as: foresight (FS); commitment (CT); and opportunity (OP). To implement these measures empirically, we average their values in Equation (10) across periods. The estimate for OP is the same for each measure, i.e., the RLM or the WCM. However, each measure will have its own values for FS and CT. We denote the estimated components of the RLM as FS_R and CT_R, while FS_C and CT_C are estimates of the components of the WCM.[15]

Foresight (FS) is our measure of "predictive" ability. Empirically, foresight is the correlation between weight differences from a benchmark and returns for any period t, giving it a value between 0 and 1. Over the measurement interval, Foresight measures the propensity of the portfolio manager to raise (lower) the weights relative to the benchmark weights on those assets that subsequently have the higher (lower) returns relative to the other assets in the portfolio. Ceteris paribus, the higher the value of FS, the more the original portfolio outperforms the benchmark.

[14] Note that for both measures (RLM and WCM), we can decompose the benchmark outperformance into these three components. Recall that for the RLM, we use the randomized weights to create the benchmark weights. For the WCM, we use the lagged weights as the benchmark weights.

[15] With respect to the WCM, although the underlying distributions are likely non-normal, we are using means and invoking the central limit theorem.

Commitment (CT) is our measure of "aggressiveness." In our formulation, commitment acts as a multiplier for foresight. Empirically, commitment is the cross-sectional standard deviation of the weight differences made to the portfolio at any time t. CT measures the propensity of the manager of the portfolio to make large (or small) changes to the portfolio weights relative to the benchmark weights. The higher the value of CT, the larger is the magnitude of the differences of the weights from the benchmark over the measurement interval. For example, if the original portfolio weights are equal to the benchmark weights, then the cross-sectional standard deviation will be zero. If the weights on the assets are very different from the benchmark, then the cross-sectional standard deviation will be large. As long as FS is positive, the higher the value of CT, the larger is the amount by which the original portfolio outperforms the benchmark.

Opportunity (OP) is our measure of the extent to which returns over a given period presented a favorable environment for a manager to generate outperformance. Empirically, opportunity is the cross-sectional standard deviation of the returns for any period t. OP measures the average cross-sectional variation in the returns over the measurement interval. Intuitively, if returns to all the assets are the same in every period, the cross-sectional standard deviation of returns will be zero. Consequently, there is no opportunity to alter portfolio returns by changing the constituent weights. However, if the returns to the assets are vastly different for each period t, the cross-sectional standard deviation of returns will be large. In this case, managers have a greater opportunity to influence portfolio return by changing relative weights.

Opportunity is also a multiplier for Foresight. Although our Opportunity measure is constant across portfolios for a given measurement interval, it can explain why a manager with a fixed skill level is more successful at outperforming the benchmark in one period versus another. All else equal between two measurement periods, if OP is higher in measurement period A than it is in measurement period B, then a manager has a better chance of outperforming the benchmark in measurement period A.

3. A Weight-Generating Process Containing Foresight and Commitment

3.1. Overview

To study how well our measures perform and how foresight and commitment affect a manager's ability to generate outperformance, we need a weight-generating process that contains foresight, commitment, and their interaction. The weight-generating process we construct is simple and straightforward.[16]

We call our weight-generating process an "active buy-and-hold" strategy. The weights generated in this process simulate a strategy wherein a manager makes incremental changes to portfolio weights from their current level. Suppose a manager believes that equity returns will be higher than those on bonds and cash. In response, the manager increases the allocation to equity from its current level. If a manager uses the "active buy-and-hold" weight-generating process, portfolio weights can drift substantially away from their initial levels.

3.2. Generating Weights for the Active Buy-and-Hold Strategy

To generate the "active buy-and-hold" weights, we modify the buy-and-hold weight formula to create an active strategy that contains varying levels of foresight and commitment. In a true passive buy-and-hold (i.e., no active portfolio weight changes), the portfolio weight for asset j at the beginning of period t+1 can be expressed as:

$$w_{j,t+1} = w_{j,t}\left(\frac{1+r_{j,t}}{1+r_{p,t}}\right) \qquad (12).$$

To create active weights that reflect foresight and commitment, we modify Equation (12) in a manner similar to Ferson and Khang (2002). Ferson and Khang (2002) assume uninformed investors form expected returns using only publicly available information. We assume uninformed investors form expected returns using only the mean and standard deviation of returns for each asset class calculated on a rolling basis over the preceding five years. Specifically, we introduce this information set as a normally distributed random error term whose mean is the average return over the trailing five years and whose standard

[16]We realize that any created weight-generating process will be somewhat arbitrary and unlikely to capture the wide variety of possibilities for how a manager may alter his portfolio. However, our created weight-generating process allows us to demonstrate important features of our measures and their component parts.

deviation is equal to the standard deviation of returns over the trailing five years. We also add a term that allows us to adjust for investor commitment.

Starting with an initial endowment of weights, $w_{j,0}$, and allowing for periods of minimum weight allocations, the weights for asset j evolve each period as follows:

$$w_{j,t+1} = max\left[w_{j,min}, w_{j,t}\left(\frac{1+\gamma(\rho r_{j,t+1}+(1-\rho)\varepsilon_{j,t+1})}{1+\gamma(\rho r_{p,t+1}+(1-\rho)\varepsilon_{p,t+1})}\right)\right] \quad (13).$$

Equation (13) results from modifying Equation (12) to allow the weights in asset j to "predict" future relative returns. That is, at the beginning of each period, weights will increase on assets whose future returns are forecasted to be higher than the portfolio's future return and weights will decrease on assets whose future returns are forecasted to be lower than the portfolio's future return. Effectively, this process changes all asset weights in a way that results in a higher portfolio return than: 1) Portfolios with constant weight levels, or 2) Portfolios with weights that drift through time in a passive buy-and-hold fashion. In our empirical tests, we place a minimum weight value of zero for all assets. When the process attempts to set an asset weight of zero, we set the weight to zero and reduce the allocation equally between the remaining assets.

To introduce varying levels of foresight, or "predictive" ability, we add the parameter ρ. In Equation (13), when $\rho = 1$, the weight evolution process completely anticipates future returns. When $\rho = 0$, the random error term, ε, completely drives the evolution. When $\rho = 0$, the evolution of the weights does not use information about future returns. However, the weights do contain historical information because the random error term uses the asset's trailing five-year historical mean and standard deviation. Values of ρ between zero and one generate beginning-of-period weights partially driven by future returns and partially driven by the random error term.[17]

[18]Even though the random error term has no information about future asset returns, the random error term should still be a "reasonable" return forecast. When $\rho = 1$, our weight generating process uses a perfect forecast of next period's return to adjust the beginning-of-period weight. When $\rho = 0$, our weight generating process uses a forecast of next period's return containing no information about the future return. However, the random forecast is "reasonable" in that the forecast for any period is a draw from a normal distribution whose mean and

To introduce relative commitment, we add the parameter γ. This parameter controls the level of aggressiveness embedded in the weights. We construct γ to accomplish two things. First, although not obvious, using γ in the way we do results in weights that always sum to one. Second, the value of γ controls the amount by which the weights change for any given set of forecasted returns and ρ. So, given a set of forecasted returns and ρ, larger values of γ result in larger weight changes. Note that this adjustment for commitment is not an absolute one. Rather, it is relative because commitment is also a function of returns and ρ.[18]

Our weight generating process, which we call the "Active Buy-and-Hold" strategy, has four attractive features. First, the process allows us to capture varying levels of foresight in an intuitive manner. Weights will increase for assets with higher relative return forecasts, while weights will decrease for assets with lower relative return forecasts. Second, given a level of foresight, our process allows for commitment level adjustments. Third, the weights generated by this process sum to one at the beginning of each period. Finally, this weight-generating process is largely consistent with an investor exhibiting exponential utility, where the weights are positively and linearly related to expected returns. That is, the "manager" generating these returns has a primary desire for higher returns.

4. Empirical Results

4.1 Returns and Weights

We can apply our performance measures to any portfolio for which data is available for beginning-of-period asset weights and subsequent asset returns. In our tests of how well our performance measures work, we use a three-asset portfolio consisting of the S&P 500, US 10-year Treasury bonds, and LIBOR. We generate beginning-of-period weights using Equation (13).

---

standard deviation are equal to the mean and standard deviation of the asset's returns over the previous five years.

[18] Assets with larger magnitude forecasted returns will experience larger magnitude weight changes. In other words, the hypothetical manager will raise his weights by more on equity if he believes equity will return 10% more than bonds and cash than he would if he believes equity will return only 1% more than bonds and cash. Our Foresight parameter, ρ, is also a factor because it determines how much of the forecasted return is used to adjust the weights.

We calculate returns using indexes from Global Financial Data (www.globalfinancialdata.com). Our measurement interval spans 29 years, 1980 through 2008.

To provide a "buy-and-hold" comparison, in Figure I we show the value of $1 million invested from the beginning of January 1980 through December 2008 for the three asset classes as well as a blended portfolio. The blended portfolio begins with a 60-30-10 percent allocation to equity, bonds, and cash. The motivation for this portfolio blend stems from previous research using this asset allocation [Arshanapalli, et al. (2001) and Brinson et. al. (1986, 1991)].

In this paper, we analyze two measurement intervals. One measurement interval spans the entire 348-month sample. The other measurement interval begins in April 1982 and ends in February 1989, the period for which we have data for weight recommendations from eight Investment Houses. For both measurement intervals, we analyze portfolio performance using weights generated with the weight-generating process described in Section 3.

Table I contains summary monthly return statistics by asset class for each measurement interval. In both intervals, equity returns have the highest average return and standard deviation, and cash returns have the lowest return and standard deviation.

Table II shows the average weights for equity, bonds, and cash across different combinations of Foresight (ρ) and Commitment (γ) for the "active buy-and-hold weight" portfolios. We present results for four levels of Foresight ($\rho = 0.0, 0.1, 0.2, 0.3$) and three levels of Commitment ($\gamma = 1, 2, 3$). As shown in Tables III and IV, these levels of Foresight and Commitment are sufficient for our measures to detect performance. In Table II, we also present the annualized geometric mean return and standard deviations over the two holding periods.

The generated weights behave largely as designed. For example, holding relative commitment, γ, constant, the average equity weight rises as foresight, ρ, rises. This behavior occurs because equity has the highest average return over the each measurement interval. This fact means that more foresight (a higher ρ) will, on average, allocate more weight to equity. Similarly, cash allocations fall as foresight increases (i.e., as ρ increases). In addition, for a given level of foresight (ρ), the standard deviation of the weights for each asset class increases as relative commitment (γ) values increase from one to three.

We can see how Commitment (γ) multiplies Foresight (ρ) through the portfolio returns shown in Table II. For example, in Panel A, for a given level of ρ, portfolio returns increase by ten to twenty basis points as γ increases by one.

4.2 Using the RLM to Judge the Performance of the "Active Buy-and-Hold" Portfolios

In Table III, we present the results of using the RLM to judge the performance of the "Active Buy-and-Hold" portfolios for two measurement intervals. The measurement intervals are January 1980 through December 2008, and April 1982 through February 1989. In addition, we present the results for the shuffled-weight versions of the FS_R, CT_R, and OP measures for twelve combinations of Foresight, ρ, and relative Commitment, γ.

As shown in Panel A, overall the annualized RLM behaves as designed. For a given level of γ, as ρ increases, the RLM increases. For example, consider the $\gamma = 1$ row in the first measurement interval. When $\rho = 0.10$, the estimated value for annualized RLM is 0.0240 with a t-statistic of 0.73. In this case, we conclude that the "manager" of this portfolio exhibits no outperformance. When $\rho = 0.20$, however, the RLM is 0.0978 with a t-statistic of 3.36. In this case, the results indicate a significant annualized outperformance of about twelve basis points. A similar pattern exists for the $\gamma = 2$ and $\gamma = 3$ rows. For both measurement intervals, we see a consistent result for the RLM for all rows in both measurement intervals. The consistent result is that the RLM is significantly different from zero for $\rho \geq 0.20$.

For the RLM, we also include the number of times (out of 10,000) that the returns from the "Active Buy-and-Hold" portfolio exceed the returns from the randomized benchmark portfolio. In the first measurement interval, when $\rho \geq 0.20$, the return from the "Active Buy-and-Hold" portfolio exceeds the returns from the shuffled benchmark portfolio more than 9,500 times out of 10,000 trials. The RLM is also significant at these levels of ρ. In the second measurement interval, the results are consistent. When the return from the "Active Buy-and-Hold" portfolio exceeds the returns from the shuffled benchmark portfolio more than 9,500 times, the RLM is significantly different from zero.

In Panel B, we present estimates of the Foresight measure, FS_R. For both measurement intervals, for a given level of γ, increases in ρ result in an increase in FS_R. For example, consider

the first measurement interval. When γ = 1 and ρ = 0.10, the FS_R measure is 0.0600 with t-statistic of 0.89. We conclude that there is no measureable foresight for this combination of γ and ρ. However, when γ = 1 and ρ = 0.20, the FS_R measure is 0.1812 with a t-statistic of 2.73. This result indicates a significant level of foresight. We see similar patterns when γ is 2 or 3. For the second measurement interval, the FS_R is significantly different from zero for $\rho \geq 0.30$. However, in both intervals, for a given level of ρ, FS_R is relatively stable as γ increases. This result is consistent with the conclusion that the FS_R is able to distinguish foresight from commitment.

The results for the CT_R commitment measure that we present in Panel C are conditional on a level of ρ. For a given level of ρ, CT_R increases as γ increases. For example, consider the ρ = 0.10 column in the first measurement interval. When γ = 1, the CT_C measure is 1.026 with a standard deviation of 0.029. When γ = 3, CT_R is 1.955 with a standard deviation of 0.054. We observe this pattern for all columns of ρ for both measurement intervals.

We report the standard deviation of CT_R (and CT_C in the next section). Note that, unlike for our measures of performance and foresight, we have no null hypothesis for the value of CT_R (and CT_C). However, the standard deviation of CT gives us a sense of the range of the estimate. For example, if CT for portfolio X was one with a standard deviation of one, and the CT for portfolio Y was four with a standard deviation of one, then the manager of portfolio X is less aggressive than is the manager of portfolio Y. However, if the standard deviations of the CTs were both eight, then it is not as certain that the manager of Portfolio X is less aggressive than is the manager of Portfolio Y.

4.3 Using the WCM to Judge the Performance of the "Active Buy-and-Hold" Portfolios

In Table IV, we present the results of using the WCM to judge the performance of the "Active Buy-and-Hold" portfolios for two measurement intervals. In addition, we present the results for the weight-based versions of the FS_C, CT_C, and OP measures for twelve combinations of Foresight, ρ, and Commitment, γ. The results generally mirror the results for the RLM.

As shown in Panel A, the annualized WCM also behaves as desired. For a given level of γ, as ρ increases, the WCM increases. For example, consider the γ = 1 row in the first

measurement interval. When $\rho = 0.10$, the estimated value for annualized WCM is 0.0256 with a t-statistic of 0.94. In this case, we conclude that the "manager" of this portfolio exhibits no outperformance. When $\rho = 0.20$, however, the WCM is 0.0903 with a t-statistic of 3.57. Here, the results indicate a significant annualized outperformance of nine basis points. A similar pattern exists for the $\gamma = 2$ and $\gamma = 3$ rows. For both measurement intervals, we see a consistent result for all rows in both measurement intervals. The consistent result is that the WCM is significantly different from zero for $\rho \geq 0.20$.

In Panel B, we present estimates of the Foresight measure, FS_C. For both measurement intervals, for a given level of γ, increases in ρ result in an increase in FS_C. For example, consider the first measurement interval. When $\gamma = 1$ and $\rho = 0.10$, the FS_C measure is 0.0643 with t-statistic of 1.58. We conclude that there is no measureable foresight for this combination of γ and ρ. However, when $\gamma = 1$ and $\rho = 0.20$, the FS_C measure is 0.1475 with a t-statistic of 3.60. This result indicates a significant level of foresight. We see similar patterns when γ is 2 or 3. Again, for both measurement intervals, we see a consistent result for all rows in both measurement intervals. The consistent result is that the FS_C is significantly different from zero for $\rho \geq 0.20$. In addition, for a given level of ρ, FS_C is relatively stable as γ increases. This result is consistent with the conclusion that the FS_C is able to distinguish foresight from commitment.

The results for the CT_C commitment measure that we present in Panel C behaves as desired, conditional on a level of ρ. For a given level of ρ, CT_C increases as γ increases. For example, consider the $\rho = 0.10$ column in the first measurement interval. When $\gamma = 1$, the CT_C measure is 0.715 with a standard deviation of 0.030. CT_C is 2.02 with a standard deviation of 0.086 when $\gamma = 3$. We observe this pattern for all columns of ρ for both measurement intervals.

4.4 Using the FS and CT statistics

The careful reader will note that, for a given level of γ, CT_R and CT_C decrease throughout as ρ increases. This result is an artifice of the "active buy-and-hold" strategy. We point out that we do not use CT_R and CT_C to measure γ. Rather, we use CT_R and CT_C to measure the overall magnitude of the weight changes made within the portfolio through time. For the "Active Buy-and-Hold" portfolios, weight changes are a function of forecasted returns and ρ as

well as γ. Therefore, this result emphasizes the fact that γ is a relative, not an absolute, adjustment for commitment.

To illustrate how to use our FS and CT statistics, consider two "Active Buy-and-Hold" portfolios from Table III found in the first measurement interval.[19] We creatively denote these two portfolios as A and B. Portfolio A has a $\rho = 0.20$ and $\gamma = 3.0$. Portfolio B has $\rho = 0.30$ and $\gamma = 2.0$. Portfolio A generates a significant RLM = 0.2718, an FS_R = 0.1772, and a CT_R = 2.444. Portfolio B generates a significant RLM = 0.3066, an FS_R = 0.3060, and a CT_R = 1.480. Both portfolios outperform the shuffled benchmark portfolio by about 35 basis points, on average. The FS_R and CT_R provide insight into how the managers generated their 30 basis point RLMs, or outperformance. When comparing FS_R statistics, the manager of Portfolio B generates outperformance with more foresight. In addition, based on the CT_R values, the manager of Portfolio B generates the outperformance level with less aggressive moves than those made by the manager of Portfolio A.

In Table III, the benchmark is the lagged weight plus a random, but actual, change. The FS_R is informative because it estimates the manager's "predictive" ability. Because the CT_R is a multiplier that magnifies the effects of the FS_R, the CT_R will contribute to outperformance only when FS_{CR} is positive. Therefore, given a level of CT_R, higher values of FS_R dominate lower values of FS_R because FS_R reflects the managers "predictive" ability.

4.4.1 Using FS and CT statistics with Investment Policy Benchmarks

In implementing our measures, we use the portfolio weights lagged one period as benchmark weights. That is, in Equation (2), $w^b_{j,t}$, represents a generic benchmark weight. In the WCM, we define $w^b_{j,t}$ as the portfolio weight lagged one period. In the RLM, we define $w^b_{j,t}$ the portfolio weight as the lagged weight plus a random, but actual, change.

However, suppose a client tells the manager that the investment policy benchmark is 50% S&P 500 Index, 30% Lehman Aggregate Bond Index, and 20% T-bills. Therefore, the benchmark weights, $w^b_{j,t,}$ represent the 50-30-20 investment policy benchmark. If the client is stern, the client will include deviations from the 50-30-20 policy benchmark when judging

[19]This example is not limited to the RLM, FS_R, and CT_R statistics. A similar analysis holds for pairs of portfolios using the WCM, FS_C, and CT_C statistics found in Table IV.

performance. In this case, FS is a measure of how much return the manager adds by deviating from the 50-30-20 policy benchmark. The CT measures how large the manager's deviations were from the 50-30-20 policy benchmark. The FS clearly indicates skill, while the CT represents higher aggression, or tracking error, relative to the investment policy benchmark.

Given similar measures of outperformance by two managers, this client will likely desire the manager with the higher FS value and lower CT value. This manager seemingly had information, but did not deviate much from the investment policy benchmark. By contrast, the other manager did not appear to have as much information, but made more aggressive moves that made up for the relatively lower level of information.

4.4.2. Portfolio Scoring

The discussion above suggests that comparisons between portfolios should be really be made within portfolio groups formed on CT. That is, a CT-based scoring mechanism would group portfolios into CT categories in a fashion similar to bond ratings. For example, in Table IV, for the second measurement interval, CT_C values fall into three categories. Portfolios with $0 < CT_C \leq 1$ could be scored as "Conservative"; The "Moderate" category is for portfolios where $1 < CT_C \leq 2$; "Highly Aggressive" portfolios are portfolios where $2 < CT_C$. Moreover, the standard deviations of CT_R and CT_C are quite low. These values, especially where $\gamma \geq 2$, reject a null hypothesis that CT values are equal across the columns of ρ. This rejection buttresses the case for a CT-based scoring mechanism because we can be confident that the CT values differ.

Within each category, comparing FS values results in portfolio performance rankings. Consider two portfolios, called C and D, found in Panel C of Table IV for the second measurement interval. Portfolio C has a $\rho = 0.10$ and $\gamma = 3.0$. Portfolio D has $\rho = 0.20$ and $\gamma = 3.0$. Portfolio C generates an insignificant WCM = 0.200 and a $CT_C = 1.9913$. Portfolio D generates a significant WCM = 0.3650, and a $CT_R = 1.8919$. Both portfolios fall into the "Moderate" category, but Portfolio D significantly outperforms Portfolio C.

4.4.3. Portfolio Scoring Interpretations

Differences between categories of the CT_R and the CT_C have a straightforward interpretation. Because the cross-sectional mean weight change at any time t is zero, the mean of $\widetilde{w}_{j,t} - \widetilde{w}^{b}_{j,t}$ is

also zero. Thus, the CT is the time-series mean of the standard deviation of the cross-sectional weight differences, $\widetilde{w}_{j,t} - \widetilde{w}^{b}_{j,t}$, where, at time t, $\widetilde{w}_{j,t}$ is the weight of asset j and $\widetilde{w}^{b}_{j,t}$, is the benchmark weight for asset j. Thus, CT is an average across the time periods of the dispersion of weight deviations from the benchmark at any given time t.

An intuitive sense of CT can be gleaned if we assume $\widetilde{w}_{j,t} - \widetilde{w}^{b}_{j,t}$ is normally distributed. For example, consider the case of CT = 2. This CT level implies that 67% of the time, the weight change from time t-1 to t was within ±2%, or ±1 standard deviation (e.g., say, for equity 60% to 62% or 60% to 58%). This CT level also implies that 95% of the time, the weight change was within ±4%, or ±2 standard deviations (e.g., say, for equity 60% to 64% or 60% to 56%).

If $\widetilde{w}^{b}_{j,t}$ represents an investment policy benchmark, then 67% of the time, the deviation from the investment policy benchmark weight was within ±2%. Suppose the equity weight benchmark is 50% via the investment policy. If CT is 2%, the manager spent two-thirds of the measurement interval with an equity weight between 48% and 52%.

5. The Performance of the Investment House Recommendations

In this section, we apply the RLM and WCM performance measures to a set of actual asset allocation recommendations. These monthly recommendations are from Bange and Miller (2004), who describe the recommendations fully. Briefly, these recommendations are published surveys of a group of Investment Houses. *The Financial Report*, a confidential newsletter purchased by *The Economist* in 1989, published these surveys. In the surveys, money managers in different countries provide asset allocation recommendations among equity, bonds, and cash for an investor who has "no existing investments, no overriding currency considerations, and the investment objective of long-term capital growth."

Several recent studies analyze the performance of these recommendations. Bange and Miller (2004) test whether past returns influence modifications to weight recommendations for equities, bonds, and cash. They find evidence of momentum trading for strategic asset allocations to equities and cash. Annaert, De Ceuster, and Van Hyfte (2005) find that the Investment Houses are not able to outperform passive benchmarks. Bange, Khang, and Miller

(2008) report scant evidence that the Investment House recommendations contain any "predictive" ability.

However, the Bange, Khang, and Miller (2008) performance measure differs in an important way from our RLM measure. Consequently, we test whether the Investment Houses fare better against our RLM and WCM performance measures. In addition, we can decompose our performance measures to examine Investment House Foresight and Commitment. We use a measurement interval spanning April 1982 through February 1989. As shown in Figure 1 of Bange, Khang, and Miller (2008), this interval is the longest continuous sample period with the greatest number of Investment Houses.

5.1 Baseline Comparison

The "Active Buy-and-Hold" strategy provides a baseline comparison to examine the performance of the returns from the Investment House weight recommendations to equity, bonds, and cash. Tables III and IV contain results for the "original" portfolios that follow this strategy. In each table, we show results for twelve combinations of ρ and γ over the April 1982 to February 1989 holding period.

As shown in Table III, the RLM is positive when $\rho \geq 0.20$. For example, when $\gamma = 1$ and $\rho = 0.20$ the level of outperformance is about 18 basis points. The RLM is significantly different from zero at the 5% level. In addition, the RLM is positive in 9,875 out of 10,000 times. Table IV contains results using the annualized WCM. Similarly, the WCM measure indicates no outperformance unless $\rho \geq 0.20$. When $\rho \geq 0.20$, values of WCM justify rejecting the null hypothesis of no outperformance by the original portfolio.

Table III also contains results for the FS_R and CT_R, while results for the FS_C and CT_C appear in Table IV. Conditional on the value for γ, the FS_R and the FS_C increases as ρ increases and are significantly different from zero when $\rho \geq 0.20$.

For each ρ and γ combination in Tables III and IV, each corresponding value for the RLM and WCM is higher in the second measurement interval. Panel D in Table IV contain the estimated measure of Opportunity, OP. In the January 1980 through December 2008 measurement interval, OP is 28.98 with a standard deviation of 1.14. In the April 1982 through

February 1989 measurement interval, OP is 32.14, with a standard deviation of 2.70. The OP measure we calculate is a sample average of the time t cross-sectional standard deviation of the returns to the assets in the portfolio. A t-test for differences in the means of the observed OP values results in a t-stat statistic of 10.44. Therefore, the period over which the Investment Houses made their weight recommendations contains greater opportunity for profit than in the overall sample period.

5.2 Investment House Performance

In Table V, we present the performance results for the Investment Houses. For each House, we present the annual return and standard deviation, the number of times out of 10,000 that the House generated a positive RLM, the annualized RLM and WCM and their t-statistics, FS_R and FS_C with their t-statistics, and the CT_R and CT_C with their standard deviations. For example, UBS Phillips and Drew had an annual return of 18.26%. Their RLM was positive 4,012 times out of 10,000 trials with an annualized RLM of -0.052% and a t-statistic of -0.263. UBS Phillips and Drew generated an FS_R of -0.022 with a t-statistic of -0.363, and a CT_R of 2.29. The annualized WCM for UBS Phillips and Drew is -0.08% with a t-statistic of -0.60. The FS_C for UBS Phillips and Drew is -0.007 with a t-statistic of -0.17, and their CT_C is 1.70.

Examining the results in Table V, one sees little evidence that the portfolios using the Investment House Recommendations outperformed based on the measures. Overall, the RLM and the WCM judged the recommended portfolios insignificantly different from zero in all cases. Upon closer examination, no Investment House generated a significant FS_R statistic. Their insignificant FS_R value suggests, perhaps, that these Houses tried but failed to generate performance by making aggressive changes to their portfolios. The values of CT_R support this conjecture. The lowest CT_R value for any House is 2.29. The highest CT_R, 6.22, results from the recommendations made by Capital House. This CT_R level greatly exceeds any CT_R level in Table III.

Finally, it is instructive to look at how a CT-based scoring mechanism would group the House-recommended portfolios into CT categories. The CT_R and the CT_C rankings have a correlation coefficient of 0.83. Moreover, the CT_R and the CT_C values place the Investment Houses into three invariant groups. Using either the CT_R or the CT_C values, the "Conservative

Group" consists of All Houses, USB Phillips and Drew, Daiwa Europe, and Anonymous Two. Bank Julius Baer, Scudder Stevens Clark, Anonymous One, and Brown Brothers Harriman comprise the "Moderate Group," although one could quibble about cut-off values. Regardless, Capital House clearly has the most aggressive CT_R and CT_C values.[20]

6. Conclusion

The existence of wide-ranging investment styles and their associated ex-ante risk often thwart attempts to measure subsequent performance of asset allocation recommendations. In this paper, we create two portfolio performance measures that match investment style and risk. One measure uses a modified version of the randomized benchmark introduced by Bange, Khang, and Miller (2008). The second measure is a modified version of the unconditional weight-based measure of Ferson and Khang (2002) and Grinblatt and Titman (1993).

Our methods create, versus pre-specify, a benchmark portfolio that shares a similar style and, perhaps ever changing, risk profile with the measured portfolio. Our measures capture the difference between portfolio returns and returns of the benchmark portfolio. We decompose outperformance into two separate, intuitive components and extend our measures to capture them. The two components we investigate are: 1) Foresight, i.e., how well can managers forecast relative returns given their information set? and 2) Commitment, i.e., how aggressively do managers act on their foresight, or, at minimum, change their weights?

Using simulated portfolio weight sets endowed with varied combinations of foresight and commitment, we find that both performance measures have power to detect performance. Our foresight measure estimates the amount of outperformance resulting from the tendency of a manager to increase (decrease) the weight on an asset above the benchmark weight when future returns are higher (lower) relative to the other assets.

We find that our commitment measure captures the magnitude of weight differences from the benchmark with enough precision to allow for the creation of a scoring mechanism that groups portfolios by commitment, and then ranks portfolios based on foresight. This

[20] The high CT_R and CT_C measures for the Houses is consistent with the (unreported) fact that the Houses made larger, though less frequent, changes to their weights relative to the "Active Buy-and-Hold" portfolios.

scoring mechanism could have important implications for investors who rely on basic asset allocation advice because one could use our measures to score active investment managers by commitment and foresight.

We apply our measures empirically by reexamining the performance of a panel of Investment House asset allocation recommendations presented in Bange and Miller (2004). Consistent with subsequent performance tests by Bange, Khang, and Miller (2008), we find no evidence that the returns from the portfolios formed by these recommendations exceed benchmark returns. We also score the Investment Houses for foresight and commitment. We find no foresight, but we document considerable variations in aggression.

In our simulated and empirical investigations, we focus on portfolio weight allocations across indexes for equity, bonds, and cash. However, it is possible to generalize our technique to many asset classes. The only inputs needed for our measures are beginning of period portfolio weights and subsequent asset-class returns. Therefore, one could use our measures for a wide variety of previously hard to measure investment styles, like those of hedge funds and other alternative investment managers. In addition, one could apply these measures to market timers and sector rotators.

An ideal property of a portfolio performance measure is that it should be able to rank managers by their proprietary skills that others cannot replicate through simple strategies [Wermers, 2006]. The motivation underpinning our measures is the desire to create a method that ranks managers by their total skill—proprietary and mechanical. If a manager fails to outperform via our measures, then the manager's performance is consistent with the performance of a manager who is not employing any mechanical or proprietary strategies. If our measures judge a manager's performance as significant, one could perform further tests, like those of Barras, Scaillet, and Wermers (2009) or Ferson and Khang (2002), to examine other aspects of the nature of the excess portfolio returns.

References

Andrews, D. W. K., I. Lee, and W. Ploberger, 1996, "Optimal changepoint tests for normal linear regression," *Journal of Econometrics* 70, 9-38.

Annaert, J., De Ceuster, M.J.K., Van Hyfte, W., 2005, "The value of asset allocation advice: Evidence from *The Economist's* quarterly poll," *Journal of Banking and Finance* 29, 661-680.

Arshanapalli, B., Coggin, D., and W. Nelson, 2001, "Is fixed-weight allocation really better?," J*ournal of Portfolio Management* 27, 27-34.

Bange, M., and T. Miller, 2004, "Return Momentum and Global Portfolio Allocations," *Journal of Empirical Finance,*11, 429–459.

Bange, M., Khang, K., and T. Miller, 2008, "Benchmarking the performance of recommended allocations to equities, bonds, and cash by international investment houses," *Journal of Empirical Finance* 15, 363-386.

Barras, L., O. Scaillet, and R. Wermers, 2009, "False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas," *Journal of Finance,* forthcoming.

Blake, D., Lehmann, B., Timmermann, A., 1999, "Asset allocation dynamics and pension fund performance," *Journal of Business* 72, 429-461.

Bollen, N. P., and Whaley, R. E., 2009. "Hedge Fund Risk Dynamics: Implications for Performance Appraisal," *Journal of Finance*, forthcoming.

Brinson, G., Hood, L., and G. Beebower, 1986, "Determinants of portfolio performance, " *Financial Analysts Journal* July/August, 39-44.

Brinson, G., Singer, B., and G. Beebower, 1991, "Determinants of portfolio performance II: an update," *Financial Analysts Journal* May/June, 40-48.

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

Copeland, T. and D. Mayers, 1982, "The *Value-Line* enigma (1965-1978): A case study of performance evaluation issues," *Journal of Financial Economics* 10, 289—322.

Daniel, K. and S. Titman, 1997, "Evidence on the characteristics of cross-sectional variation in common stock returns," *Journal of Finance* 52, 1—33.

Fama, E., and K. French, 1993, “Common Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics* 33, pp. 3-56.

Ferson, W., and K. Khang, 2002, “Conditional performance measurement using portfolio weights: evidence for pension funds,” *Journal of Financial Economics* 65, 249-282.

Ferson, W., and R. W. Schadt, 1996, “Measuring Fund Strategy and Performance in Changing Economic Conditions,” *Journal of Finance* 51, pp. 425-461.

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

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

Grinblatt, M., and S. Titman, 1993, “Performance measurement without benchmarks: an examination of mutual fund returns,” *Journal of Business* 60, 97-112.

Grinblatt, M. and S. Titman, 1989, “Mutual fund performance: An analysis of quarterly portfolio holdings,” *Journal of Business* 62, 394—415.

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, 513-533.

Ivković, Z., Sialm, C., and Weisbenner, S., 2008,” Portfolio concentration and the performance of individual investors, “ *Journal of Financial and Quantitative Analysis* 43, 613-656.

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

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

Kacperczyk M., Sialm, C., and Zheng, L., 2005, “On the Industry Concentration of Actively Managed Equity Mutual Funds,” *Journal of Finance* 60, 1983-2011.

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.

Roll, R., 1992, “A mean/variance analysis of tracking error,” *Journal of Portfolio Management* 18. 13—23.

Roll, R., 1978, “Ambiguity when Performance is Measured by the Securities Market Line,” *Journal of Finance* 33, pp. 1051-1069.

Roll, R., 1977, “A Critique of the Asset Pricing Theory’s Tests. Part I: On Past and Potential Testability of the Theory,” *Journal of Financial Economics* 4, pp. 129-176.

Sharpe, W. F., 1992, “Asset allocation: Management style and performance measurement, *Journal of Portfolio Management,* pp. 7-19.

Sharpe, W. F., 1964, “Capital asset prices: A theory of market equilibrium under conditions of risk,” *Journal of Finance* 19, 425-442.

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

Wermers, R., 2006, “Performance Evaulation with Portfolio Holdings Information,” *North American Journal of Economics and Finance* 17, pp. 207-230.

Wermers, R., 2000, “Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transaction Costs, and Expenses,” *Journal of Finance* 55, pp. 1655-1703.

Figure 1: Value of $1 million Invested from January 1980 - December 2008

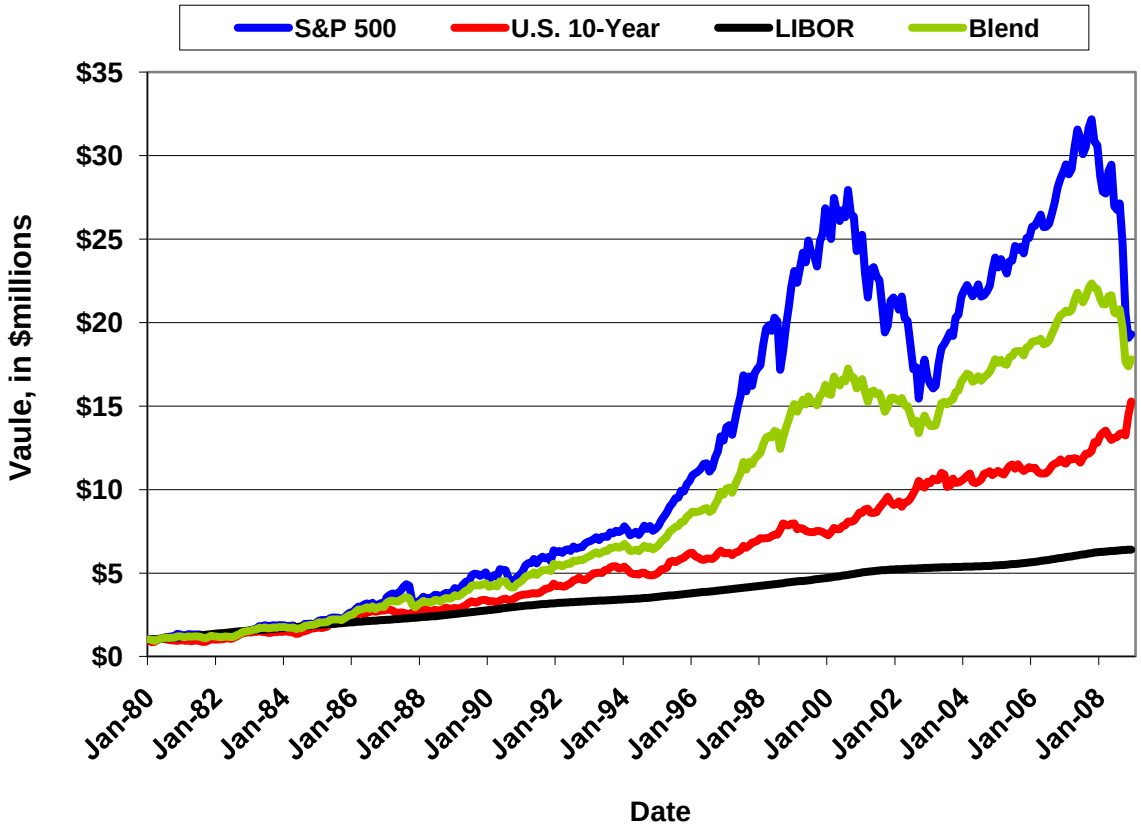

Table I

Asset Class Returns

The table presents summary statistics for the monthly asset class returns over two sample periods. The table includes the mean, standard deviation, maximum, minimum, and number of observations for the S&P 500, the US 10-year Treasury bond, and LIBOR. The table also includes the number of times the monthly return was less than the time series average LIBOR and the number of times the monthly return was negative.

Panel A. January 1980 through December 2008

	S&P 500	US 10-Yr.	LIBOR
Mean Monthly Return	1.0%	0.8%	0.5%
Standard Deviation	4.4%	2.8%	0.3%
Maximum	13.5%	14.3%	1.7%
Minimum	-21.5%	-9.0%	0.1%
Observations	348	348	348
Mo. Return < Avg. LIBOR	145	165	211
Mo. Return < 0	128	123	0

Panel B. April 1982 through February 1989

	S&P 500	US 10-Yr.	LIBOR
Mean Monthly Return	1.6%	1.3%	0.7%
Standard Deviation	4.9%	3.3%	0.2%
Maximum	13.5%	8.9%	1.3%
Minimum	-21.5%	-5.7%	0.5%
Observations	83	83	83
Mo. Return < Avg. LIBOR	33	39	48
Mo. Return < 0	29	31	0

Table II

Average Weight Allocations and Returns for the "Active Buy-and-hold" Portfolios

The table presents summary statistics for the weight allocations of the "active buy-and-hold" portfolios for two measurement intervals for various levels of Foresight (ρ) and Commitment (γ). We generate the weights in this table using Equation (13). This table includes the mean and standard deviation for the portfolio weight allocations, in percent, to equity, bonds, and cash. The S&P 500, the US 10-year Treasury bond, and LIBOR represent these three asset classes. We use Foresight (ρ) levels that range from 0 (representing no foresight) to 0.30. We use Commitment (γ) levels with values of 1 (representing no Foresight multiplier), 2, and 3. Returns are annualized, in percent.

N = 348	Foresight, (ρ):	$\rho = 0.00$		$\rho = 0.10$		$\rho = 0.20$		$\rho = 0.30$	
Commitment, γ		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
$\gamma = 1$	S&P 500	65.0	3.8	66.3	4.0	67.4	4.2	68.5	4.6
	US 10-Year	27.5	2.9	26.6	2.6	25.8	2.6	25.0	2.7
	LIBOR	7.5	1.8	7.1	2.0	6.8	2.1	6.6	2.2
	Returns	*10.53*	*10.60*	*10.56*	*10.76*	*10.60*	*10.91*	*10.63*	*11.04*
$\gamma = 2$	S&P 500	65.0	7.0	68.3	6.5	71.0	6.7	73.2	7.1
	US 10-Year	27.7	6.3	25.3	5.2	23.3	4.7	21.6	4.5
	LIBOR	7.3	2.7	6.4	2.9	5.7	3.1	5.2	3.2
	Returns	*10.67*	*10.65*	*10.73*	*11.04*	*10.75*	*11.38*	*10.82*	*11.66*
$\gamma = 3$	S&P 500	60.4	11.6	66.9	9.4	71.8	8.6	75.4	8.7
	US 10-Year	30.7	11.4	26.3	8.6	22.6	6.8	19.8	6.0
	LIBOR	8.8	3.4	6.8	3.6	5.6	3.8	4.8	3.9
	Returns	*10.79*	*10.26*	*10.94*	*10.93*	*11.01*	*11.50*	*11.05*	*11.94*

Panel A. Average Simulation Weights, in Percent, January 1980 through December 2008

N = 83	Foresight, (ρ):	$\rho = 0.00$		$\rho = 0.10$		$\rho = 0.20$		$\rho = 0.30$	
Commitment, γ		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
$\gamma = 1$	S&P 500	66.8	2.7	66.6	2.6	66.3	2.6	66.0	2.5
	US 10-Year	23.9	2.1	24.4	1.9	24.9	1.8	25.5	1.7
	LIBOR	9.3	0.9	9.0	0.9	8.7	0.9	8.5	1.0
	Returns	*17.79*	*11.70*	*18.28*	*12.48*	*18.31*	*12.47*	*18.34*	*12.45*
$\gamma = 2$	S&P 500	72.3	4.6	72.0	4.5	71.6	4.4	71.1	4.3
	US 10-Year	18.6	3.5	19.5	3.3	20.5	3.1	21.4	3.0
	LIBOR	9.1	1.6	8.5	1.5	7.9	1.5	7.5	1.5
	Returns	*18.58*	*13.18*	*18.65*	*13.17*	*18.72*	*13.15*	*18.77*	*13.11*
$\gamma = 3$	S&P 500	76.5	5.8	76.3	5.8	75.9	5.7	75.3	5.6
	US 10-Year	14.2	4.4	15.4	4.2	16.6	4.1	17.9	3.9
	LIBOR	9.3	2.2	8.3	2.0	7.5	1.9	6.8	1.9
	Returns	*18.80*	*13.71*	*18.92*	*13.74*	*19.02*	*13.72*	*19.11*	*13.66*

Table III

RLM, FS_R, CT_R, and OP Measures for the Active Buy-and-hold Portfolios

This table presents the RLM, FS_R, CT_R, and OP measures for the "active buy-and-hold" portfolios over two measurement intervals for various levels of Foresight (ρ) and Commitment (γ). We show the method to calculate RLM in the text in Equations (4) through (7). RLM is the difference between the return on the active portfolio and the mean of 10,000 randomized benchmark returns obtained by shuffling the order of weight changes in the measurement interval. FS_R is the foresight measure, and CT_R is the commitment measure decomposed from the RLM. The Opportunity measure, OP, is constant for all Foresight (ρ) and Commitment (γ) levels in each measurement interval. We show the decomposition method in the text in Equations (10), (10a-f), and (11).

.

1. January 1980—December 2008, N=348

Panel A: RLM, z-statistic for H_0: RLM = 0, and Times portfolio beats benchmark

	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	-0.0523	0.0240	**0.0978**	**0.1692**
	(-1.41)	(0.73)	**(3.36)**	**(6.47)**
	796	7,702	**9,998**	**10,000**
$\gamma = 2$	-0.1005	0.0482	**0.1952**	**0.3066**
	(-1.39)	(0.77)	**(3.46)**	**(6.44)**
	827	7,807	**9,998**	**10,000**
$\gamma = 3$	-0.1484	0.0768	**0.2718**	**0.4354**
	(-1.37)	(0.82)	**(3.45)**	**(6.49)**
	862	7,965	**9,998**	**10,000**

Panel B: FS_R and z-statistic for H_0: $FS_R = 0$

	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	-0.0383	0.0600	**0.1812**	**0.3183**
	(-0.57)	(0.89)	**(2.73)**	**(4.86)**
$\gamma = 2$	-0.0369	0.0595	**0.1797**	**0.3060**
	(-0.55)	(0.88)	**(2.69)**	**(4.66)**
$\gamma = 3$	-0.0342	0.0632	**0.1772**	**0.3008**
	(-0.51)	(0.94)	**(2.65)**	**(4.57)**

Panel C: CT_R and standard deviation

	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	1.160	1.026	0.911	0.820
	0.032	0.029	0.025	0.023
$\gamma = 2$	2.267	1.955	1.754	1.480
	0.063	0.054	0.047	0.041
$\gamma = 3$	3.419	2.912	2.444	2.074
	0.093	0.080	0.067	0.057

2. April 1982—February 1989, N=83

Panel A: RLM, z-statistic for H_0: RLM = 0, and Times portfolio beats benchmark

	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	-0.0093	0.0753	**0.1608**	**0.2473**
	(-0.10)	(0.84)	**(1.96)**	**(3.28)**
	4,542	7,949	**9,790**	**9,999**
$\gamma = 2$	-0.0066	0.1433	**0.2961**	**0.4530**
	(-0.04)	(0.90)	**(2.02)**	**(3.30)**
	4,755	8,115	**9,817**	**9,999**
$\gamma = 3$	0.0027	0.2005	**0.4025**	**0.6136**
	(0.01)	(0.96)	**(2.06)**	**(3.32)**
	4,942	8,309	**9,835**	**9,999**

Panel B: FS_R and z-statistic for H_0: $FS_R = 0$

	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	0.0094	0.0944	0.1877	**0.3004**
	(0.07)	(0.74)	(1.49)	**(2.39)**
$\gamma = 2$	0.0100	0.0953	0.1911	**0.3054**
	(0.08)	(0.76)	(1.52)	**(2.43)**
$\gamma = 3$	0.0132	0.1003	0.2004	**0.3037**
	(0.10)	(0.80)	(1.60)	**(2.43)**

Panel C: CT_R and standard deviation

	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	1.356	1.239	1.133	1.044
	0.082	0.076	0.071	0.066
$\gamma = 2$	2.336	2.159	2.001	1.868
	0.140	0.132	0.125	0.118
$\gamma = 3$	3.001	2.797	2.625	2.486
	0.174	0.167	0.161	0.154

Table IV

WCM, FS_C, CT_C, and OP Measures for the Active Buy-and-hold Portfolios

This table presents the WCM, FS_C, CT_C, and OP measures for the "active buy-and-hold" portfolios over two measurement intervals for various levels of Foresight (ρ) and Commitment (γ). We show the method to calculate WCM in the text in Equations (8) and (9). Conceptually, this weight-based measure estimates the covariance between weight changes and subsequent returns. FS_C is the foresight measure, and CT_C is the commitment measure decomposed from the WCM. The Opportunity measure, OP, is constant for all Foresight (ρ) and Commitment (γ) levels in each measurement interval. We show the decomposition method in the text in Equations (10), (10a-f), and (11).

Panel A: WCM for annualized returns, in percent, and a t-statistic for H_0: WCM = 0

	1. January 1980—December 2008, N = 348					2. April 1982—February 1989, N=83			
	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$		$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	-0.0411	0.0256	**0.0903**	**0.1529**	$\gamma = 1$	0.0070	0.0764	**0.1470**	**0.2187**
	(-1.28)	(0.94)	**(3.57)**	**(5.79)**		(0.11)	(1.37)	**(2.51)**	**(3.03)**
$\gamma = 2$	-0.0766	0.0523	**0.2043**	**0.2772**	$\gamma = 2$	0.0221	0.1440	**0.2693**	**0.3994**
	(-1.23)	(1.00)	**(3.46)**	**(5.64)**		(0.20)	(1.45)	**(2.56)**	**(3.06)**
$\gamma = 3$	-0.1099	0.0844	**0.2526**	**0.3942**	$\gamma = 3$	0.0399	0.2010	**0.3650**	**0.5398**
	(-1.19)	(1.09)	**(3.66)**	**(5.59)**		(0.27)	(1.52)	**(2.62)**	**(3.11)**

Panel B: FS_C and a t-statistic for H_0: $FS_C = 0$

	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$		$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	-0.0327	0.0643	**0.1512**	**0.2455**	$\gamma = 1$	0.0297	0.1253	**0.1935**	**0.2488**
	(-0.79)	(1.58)	**(3.71)**	**(6.17)**		(0.36)	(1.56)	**(2.44)**	**(3.15)**
$\gamma = 2$	-0.0329	0.0628	**0.1475**	**0.2369**	$\gamma = 2$	0.0277	0.1236	**0.1944**	**0.248**
	(-0.80)	(1.53)	**(3.60)**	**(5.94)**		(0.33)	(1.52)	**(2.44)**	**(3.13)**
$\gamma = 3$	-0.0298	0.0636	**0.1452**	**0.2334**	$\gamma = 3$	0.0265	0.1210	**0.1948**	**0.2479**
	(-0.72)	(1.55)	**(3.54)**	**(5.83)**		(0.31)	(1.48)	**(2.43)**	**(3.12)**

Panel C: CT_C and standard deviation

	$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$		$\rho = 0.00$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.30$
$\gamma = 1$	0.811	0.7150	0.637	0.5748	$\gamma = 1$	0.9734	0.8917	0.8257	0.7641
	(0.033)	(0.030)	(0.026)	(0.024)		(0.07)	(0.07)	(0.06)	(0.05)
$\gamma = 2$	1.583	1.360	1.201	1.0311	$\gamma = 2$	1.6709	1.5489	1.4531	1.3633
	(0.065)	(0.057)	(0.054)	(0.044)		(0.13)	(0.12)	(0.11)	(0.10)
$\gamma = 3$	2.381	2.020	1.696	1.4348	$\gamma = 3$	2.1298	1.9913	1.8919	1.8026
	(0.099)	(0.086)	(0.065)	(0.063)		(0.16)	(0.15)	(0.14)	(0.13)

Panel D: OP w/ standard deviation

OP	28.98	OP	32.14
	1.14		2.70

Table V

Performance Measures, Foresight, and Commitment for the Investment Houses Recommendations

In this table, we present the RLM and WCM performance measures and the FS_R, CT_R. FS_C, and CT_C for Investment House recommendations. The RLM is the randomized lagged weight measure. The WCM is the weight-based cross-sectional return measure. The FS_R and FS_C are randomized and weight-based Foresight measures. The CT_R and CT_C are the randomized and weight-based Commitment measures. The first two columns of the table present the Investment House name and annualized return to a portfolio that follows the House recommendations. The next four columns are: the times out of 10,000 weight-change shuffles that the RLM is positive; the RLM with its t-statistic; the FS_R with its t-statistic; and the CT_R with its standard deviation. The last three columns are: the annualized WCM and its t-statistic; the FS_C and its t-statistic; and, the CT_C and its standard deviation.

April 1982 through February 1989 (N = 83)	Return (%) St. Dev.	Return > Shuffle	RLM T-Stat	FS_R T-Stat	CT_R St. Dev	WCM (%) T-Stat	FS_C T-Stat	CT_C St. Dev
UBS Philips and Drew	18.26	4,012	-0.052	-0.022	2.29	-0.08	-0.007	1.70
	11.89		-0.263	-0.363	0.27	-0.60	-0.170	0.37
Capital House	17.19	5,070	0.011	-0.018	6.22	0.22	-0.027	4.28
	12.68		0.018	-0.234	0.46	0.48	-0.518	0.77
Scudder Stevens Clark	18.46	5,425	0.048	-0.082	3.42	0.08	-0.017	2.32
	15.54		0.131	-0.825	0.21	0.34	-0.281	0.33
Bank Julius Baer (N = 71)	12.82	3,026	-0.211	0.025	3.19	-0.06	0.012	2.42
	9.80		-0.548	0.348	0.31	-0.20	0.206	0.42
Daiwa Europe	18.77	2,031	-0.244	-0.051	2.45	-0.12	-0.059	1.59
	14.05		-0.883	-1.245	0.37	-0.38	-1.845	0.61
Brown Brothers Harriman	17.82	9,160	0.544	0.010	3.68	0.59	-0.016	2.31
	14.11		1.424	0.139	0.31	1.22	-0.336	0.67
Anonymous One	18.79	2,627	-0.208	-0.051	3.44	-0.19	0.008	2.40
	15.19		-0.633	-0.745	0.29	-0.59	0.166	0.54
Anonymous Two	18.44	8,576	0.286	-0.006	2.76	0.19	-0.027	1.77
	15.99		1.056	-0.088	0.26	0.91	-0.614	0.42
Average Weight of All Houses	18.22	7,014	0.099	-0.042	2.17	0.10	-0.005	1.46
	13.48		0.510	-0.405	0.11	0.56	-0.078	0.19