Performance Measurement with Market and Volatility Timing and Selectivity

by
Wayne Ferson
and
Haitao Mo

First draft: September 5, 2011
This revision: June 29, 2013

ABSTRACT:

The investment performance of portfolio managers depends on market level and volatility timing as well as security selection. We develop new holdings-based performance measures that properly adjust for risk, accommodate all three components in a consistent framework and avoid strong assumptions about managers' behavior. We find that mutual funds with more active responses to volatility have better subsequent performance. Funds show greater (less) ability to time market factor levels when an investor sentiment measure is low (high). Sorting funds by factor model R-squares and other evidence confirm previous findings that the more active funds have better performance.

* Ferson is the Ivadelle and Theodore Johnson Chair of Banking and Finance and a Research Associate of the National Bureau of Economic Research, Marshall School of Business, University of Southern California, 3670 Trousdale Parkway Suite 308, Los Angeles, CA. 90089-0804, ph. (213) 740-5615, ferson@marshall.usc.edu, www-rcf.usc.edu/~ferson/. Mo is a Ph.D student in Finance, Marshall School of Business, University of Southern California, haitaomo@usc.edu. We are grateful to Fabio Moneta for help with some data, to participants at the 2012 Northern Finance meetings, the 2012 Sonoran Finance Conference, the 2012 Alberta Frontiers in Finance conference, the 2012 conference on Financial Economics, Accounting and Management, the 2013 McGill Asset Management Conference and workshops at the following Universities: Buffalo, Emory, Southern California, Texas and Texas A+M for helpful discussions, and to Yakov Amihud, Oliver Boguth, Jeff Busse, Yong Chen and Sung June Pyun for comments.

Performance Measurement with Market and Volatility Timing and Selectivity

ABSTRACT:

The investment performance of portfolio managers depends on market level and volatility timing as well as security selection. We develop new holdings-based performance measures that properly adjust for risk, accommodate all three components in a consistent framework and avoid strong assumptions about managers' behavior. We find that mutual funds with more active responses to volatility have better subsequent performance. Funds show greater (less) ability to time market factor levels when an investor sentiment measure is low (high). Sorting funds by factor model R-squares and other evidence confirm previous findings that the more active funds have better performance.

1. Introduction

Researchers' attempts to measure the investment performance of portfolio managers have long been hobbled by market timing. If fund managers attempt to trade in anticipation of market-wide factors (market timing behavior), it has been known since Grant (1977) that security selection ability is hard to measure. If managers attempt to both time the markets and pick undervalued securities, it is hard to distinguish one skill from the other. Commonly, market timing and selectivity performance are measured assuming that only one of those abilities exists. This can lead to a missing variables bias if both types of behavior are present. In addition, funds may attempt to time or react to market volatility (Busse, 1999), further complicating inferences. This paper develops new performance measures that accommodate selectivity, level timing and volatility timing in a consistent framework.

Classical measures of timing ability such as Treynor and Mazuy (1966) and Merton and Henriksson (1981) make strong assumptions, such as timing behavior of a stylized form, and/or the validity of a simple options pricing model. This leads to misspecification if the stylized assumptions are not satisfied. Our approach avoids such stylized assumptions by observing managers' behavior through the portfolio weights.

Measures of performance that attempt to accommodate market timing behavior typically model the ability to time the level of market factors, but not market volatility. Investors value market level timing because the positive covariance between a fund's market exposure and the future market return boots the expected portfolio return for a given average risk exposure. Risk-averse investors value volatility timing when funds can reduce market exposure in anticipation of higher volatility. The negative covariance between a fund's market exposure and volatility lowers the average volatility of the portfolio, and can do so without an average return penalty.[1] Busse (1999) studies volatility timing behavior in US mutual funds, and finds evidence for the behavior in funds' returns. Aragon and Martin (2012) suggest that hedge funds may actively time volatility. If both level and volatility timing behavior are present, models that leave one of them out are likely misspecified.[2]

Classical measures identify timing ability through the nonlinearity it creates in funds' returns, using stylized models of manager behavior to map the nonlinearity into managers' superior information. There are two problems with this approach. First, there are many potential sources of nonlinearity in fund returns that may be unrelated to market timing ability (see Chen Ferson and Peters (2010) for a discussion). Second, if the nonlinear timing term is left out of a returns-based performance regression, the selectivity measure is biased when the missing nonlinear term is correlated with the included linear term. By using portfolio holdings instead of reported returns, we avoid these problems. By observing fund managers' behavior through their holdings, we avoid stylized assumptions about the behavior.

---

[1] If the market volatility process is such that high unexpected volatility leads to an upward revision in market volatility, and if prices adjust downward to raise expected returns for the higher volatility, then a volatility-timing strategy will enhance portfolio average returns.

[2] Holmes and Faff (2004) apply Busse's model in Australia, and Kim and In (2012) examine Busse's model using simulations. Other studies further motivate the relevance of volatility timing. Fleming, Kirby and Ostdiek (2001), Johannes, Polson and Stroud (2002) and Han (2009) find that models attempting to predict volatility have an economically significant impact on mean-variance optimal portfolio strategies. Graham and Harvey (1996) find that dispersion in newsletters' asset allocation recommendations predicts future market volatility. Aragon and Martin (2012) find that aggregate hedge fund demand for options predicts market volatility changes. Cao, Chen, Liang and Lo (2012) consider both market and volatility timing for hedge funds, but their returns-based approach is very different

This paper contributes to a rapidly-developing literature on holdings-based performance measures, kicked off by Grinblatt and Titman (GT, 1989). Section 2 discusses the relation between returns-based and holdings-based performance measures. We also (Section 3.1) develop the relation of our measures to previous holdings-based measures, and point out sources of misspecification in those measures that our new measures avoid. The intuition for the misspecification in the earlier holdings-based measures follows from the fact that these measures examine $\mathrm{Cov}(x_t' r_{t+1})$, which denotes the sum over the securities i, of the covariances between portfolio holdings, x_{it}, and the subsequent realized excess returns, r_{it+1}.[3] However, a well-specified performance measure is based on $\mathrm{Cov}(x_t' m_{t+1} r_{t+1})$, the sum of the covariances between the portfolio holdings and the subsequent abnormal, or risk-adjusted returns, where m_{t+1} is the stochastic discount factor (SDF). We use popular linear SDFs in this paper, but the idea can be used with any SDF.

Using linear SDFs, we show that previous holdings-based performance measures leave out a volatility timing component. Boguth et al. (2011) suggest that volatility timing may impart substantial biases to estimates of alpha in other contexts. Our analysis also implies that previous holdings-based measures of selectivity leave out a second moment term, where the selectivity component of a fund's return may be correlated with the SDF. These missing terms can change the inference about funds' ability. In particular, we find that investment ability is significantly related to a fund's tendency to react to market volatility, while the selectivity measure of Grinblatt, Titman and Wermers (1997) finds no significant relation.

---

from ours.

[3] We characterize the GT measure as an estimate of the covariance. Empirically, GT estimate $E\{[x-x_{lag}]'r\}$, a weight-change measure, where the lagged weight, x_{lag} serves as a

Our measures extend the previous holdings-based performance measures in a parsimonious way. Only three parameters are needed for each mutual fund. This allows us to easily examine models with multiple benchmarks.

We implement our measures on a sample of US active, open-ended mutual funds. We find that funds with more active responses to volatility have better subsequent performance. We also find that the ability of funds to time market factor levels is correlated with a lagged measure of investor sentiment and with the aggregate flows of new money. Sorting funds by factor model R-squares confirms the findings of Amihud and Goyenko (2013) that the low R-square funds have better performance.

The rest of the paper is organized as follows. Section 2 briefly reviews returns-based versus holdings-based performance measures. Section 3 describes our models and their estimation. Section 4 describes the data, Section 5 presents the empirical results and Section 6 concludes.

2. Returns-based versus Holdings-based Performance Measures

Returns-based measures of performance typically compare the after-fund-cost returns of a fund with the returns of a fund-specific benchmark. In principle, the benchmark should be a feasible "Otherwise Equivalent" (Aragon and Ferson, 2008) alternative choice to the fund, except without the fund's skill. Most typically, risk adjustments are used to define equivalent. For example, Jensen's (1968) alpha, based on the Capital Asset Pricing Model (Sharpe, 1964), uses a fund beta-weighted average of a market index and short term cash securities as the benchmark. A fund whose average return exceeds this benchmark has a

---

proxy for the expected weight.

positive Jensen's alpha.

Since reported mutual fund returns are measured after expenses and funds' trading costs, returns-based measures are directed at what is left for investors after funds' costs. However, returns-based measures in practice typically mix the after-cost returns of funds with the before-cost returns of the benchmarks, creating an "apples to oranges" comparison. Accurate returns-based measures also require accurate measures of funds' risks. Even if the risk model could be agreed on, risk measurement can be difficult when funds engage in market timing or other activities that imply high-frequency trading and risk exposures that vary over time (e.g., Ferson and Schadt (1996), Patton and Ramadorai (2013).

Holdings-based measures, in contrast, examine the covariances between funds' holdings and the subsequent before-cost returns of the underlying securities held, looking for the ability to take positions before securities rise in value. Because the underlying assets' returns are measured before costs, holdings-based measures are better suited than returns-based measures for assessing managers' investment ability before trading costs and fees.

Holdings-based approaches avoid the problems induced by high frequency trading, but do not exploit the information in high frequency returns, which Bollen and Busse (2001) find helps to detect market timing ability. This can result in a loss of power. However, Ferson and Khang (2002) and Jiang et al. (2007) examine the power of holdings-based approaches with simulation and find that including the large amount of information in a fund's portfolio holdings more than offsets the loss of information from a single time-series of the reported fund returns. Measures using funds' holdings can be quite powerful. The small standard errors we report are consistent with better precision than returns-based measures.

Reported holdings can be subject to "window dressing" (e.g. Lakonishok, Shliefer and Vishny, 1992), where funds attempt to mislead investors at reporting dates (see Solomon, Soltes and Sosyura (2012) for a recent analysis). Such behavior could obscure truly informed trading, but backward-looking window dressing should not produce false predictive ability. Abstracting from costs, weight-based measures miss the possibility that the ability to trade at low cost or the ability to manage an efficient securities lending operation can be forms of skill. More deeply, while the returns and holdings are measured before costs, fund managers presumably determine their holdings through some optimization in a world with costs. Modeling this consideration in holdings-based measures is a good opportunity for future research.[4]

3. The Models

Consider a definition of abnormal performance, or alpha, based on the Stochastic Discount Factor:

$$\alpha_p = E(m\ r_p), \tag{1}$$

where m is the SDF and r_p is the return of the fund in excess of a short-term Treasury bill. This measure of performance goes back to Grinblatt and Titman (1989), Glosten and Jagannathan (1994) and Chen and Knez (1996) who adopt specific SDF models. Ferson and Lin (2012) argue that if m is the client's marginal rate of substitution, Equation (1) is the best

[4] We thank David Mauer for this idea.

way to specify a valid performance measure.

To see the structure of holdings-based performance measures let x_t be the vector of holdings at time t and let $r_{pt+1} = x_t' \, r_{t+1}$ be the "hypothetical" fund excess return based on the vector of future excess returns, r_{t+1}, on the securities held at time t (we now suppress the time subscripts unless required for clarity). Ferson (2012) shows that Equation (1) and simple algebra yield:

$$\alpha_p = \mathrm{Cov}(x' \, m \, r). \tag{2}$$

Equation (2) says that the right way to measure performance with holdings is the sum of the covariances between the portfolio weights and the risk-adjusted abnormal returns, m r. The classical Grinblatt and Titman (1989, 1993) measure is an estimate of Cov(x'r), which leaves out the risk adjustment, m. Ferson (2012) provides conditions under which the two approaches are equivalent, but the conditions are stringent and unlikely to be met in practice. Our measures are versions of Equation (2).

We assume the SDF is given by popular linear factor models, following Cochrane (1996), and using factors that are current standards in the fund performance literature:

$$m = a - b' r_B, \tag{3}$$

where r_B is a vector of K benchmark portfolio excess returns and a and b are market-wide parameters to be estimated. The simplest example is the CAPM, where K=1 and a broad stock market index is the benchmark. We also use the Fama-French (1996) and Carhart (1997) factors, and models that include a bond market factor. We start with equations (1) -

(3) in their simple "unconditional" form and discuss conditional versions of the model below.[5]

Consider a factor model regression for the excess returns of the N underlying securities:

$$r = a_0 + \beta\, r_B + u, \tag{4}$$

where β is the N x K matrix of regression betas and $E(ur_B)=E(u)=0$. Let the vector of idiosyncratic returns be the sum of the intercepts plus residuals: $v = a_0 + u$. A fund forms a portfolio of the N assets using weights, x, as:

$$r_p = x'r = (x'\beta)\, r_B + x'v. \tag{5}$$

In this formulation, the "cash" position invested in the short term Treasury security is $1 - \underline{1}'x$, where $\underline{1}$ is an N-vector of ones. Define $w'=x'\beta$ as the "asset allocation" weights. Our approach is to estimate these using a "bottom up" method and daily data for the underlying asset returns and benchmarks, similar to Jiang, Yao and Yu (2007) and Elton, Gruber and Blake (2010). Substituting Equation (5) into the definition of alpha we obtain:

$$\begin{aligned} \alpha_p &= a\, E(w'r_B) - b'\, E(r_B r_B'\, w) + E\{(a-b'r_B)\, x'v\}, \qquad (6) \\ &= a\, \mathrm{Cov}(w'r_B) - b'\, E\{\, [r_B r_B' - E(r_B r_B')]\, w\} + E\{(a-b'r_B)\, x'v\}. \end{aligned}$$

[5] Many interesting extensions are possible. For example, Kang (2012) includes a liquidity factor in the model and finds it useful for hedge fund performance. Including squared benchmark returns brings in skewness preference and the possibility of "skewness timing." Our approach is parsimonious enough to handle additional factors. Future research should explore these and other extensions.

The benchmarks have zero alphas in (1) by construction when (3) describes the SDF. This allows us to move between the first and second lines of Equation (6), writing the measures in terms of covariances. In our estimation scheme, described below, we use the second line of Equation (6).

The first term in Equation (6) is essentially the weight-based measure of Grinblatt and Titman (GT, 1989, 1993) applied at the "asset allocation" level. This captures market level timing through the covariance between the asset allocation weights and the subsequent benchmark returns. A fund that puts more weight in asset classes that subsequently offer unexpectedly high returns has positive level timing. The second term captures "volatility timing." A fund that puts more weight on factors whose second moments are subsequently unexpectedly low is attractive to investors who dislike second moments, so this term gets a negative coefficient, -b.

The third term of Equation (6) captures selectivity ability. This term focuses on the portfolio-weighted average of the idiosyncratic security returns, $x_t' v_{t+1}$. When this dynamic strategy is positive, or has a positive covariance with the marginal rate of substitution, it represents selectivity performance with positive value. There are two components of selectivity. The first, $aE(x'v)$, is similar to previous measures. The second, $E\{(-b'r_B)(x'v)\}$, captures the relation between a fund's dynamic combination of the stocks' idiosyncratic returns and marginal utility, as reflected in the benchmark returns, $(-b'r_B)$. This higher moment term is ignored in previous measures.[6]

---

[6] Even though v_{t+1} and r_{Bt+1} are uconditionally uncorrelated, they may be related through the information in x_t. For example, if the portfolio weight x_t has selectivity information about v_{t+1}, the product of the two is related to an average of the stocks' idiosyncratic

The sum of the three components of performance has a simple relation to a traditional regression-based alpha. If the "hypothetical" before cost excess returns, $r_p=x'r$ are used in a factor model regression on r_B, the intercept is proportional to α_p. This is because, on the assumption of a linear factor model for the SDF, the SDF alpha is proportional to the intercept in the factor model regression (e.g. Ferson, 1995). By exploiting the holdings data, x, our measures make it possible isolate the three components of performance. Unlike in classical returns-based measures, this is possible without making stylized assumptions about fund manager behavior, because we observe their behavior through the portfolio holdings.[7] (Because the hypothetical return based on the holdings does not reflect any interim trading, it avoids many of the issues discussed above, where high frequency trading biases regression-based estimates of alpha.)

Previous measures of market timing, such as the classical quadratic regression of Treynor and Mazuy (1966) are difficult to apply for more than one or two benchmarks. This is because the regression includes on the right hand side the benchmarks, the squared benchmarks, and with multiple benchmarks, the products of the benchmarks. With K benchmarks, there are $2K + (K^2 - K)/2$ fund-specific coefficients to be estimated. For example with three factors there are 9 slope coefficients plus an intercept in the regression

---

volatilities. If there is a discount rate effect for average idiosyncratic volatility, such that the volatility and the benchmark return have a negative correlation, then the higher moment effect is positive. There are of course other examples where $x_t'v_{t+1}$ contains information about r_{Bt+1}. Since the higher moment effect involves the benchmark return, there is room for debate over calling it selectivity or a form of timing. We choose the former for two reasons. First, the term involves only the idiosyncratic components of the security returns, v. Second, the first two terms in Equation (6) as specified, cleanly capture level and volatility timing.

[7] Letting $r_p = a_p + \beta_p r_B + e$, with $E(e)=0=E(er_B)$, and $m=a-b'r_B$, then $\alpha_p= E(mr_p) = E(m)\, a_p$. Thus, the hypothetical return is sufficient for estimating total performance, and the regression intercept a_p is equivalent to our total performance measure. But we cannot

for each fund. This can be a degrees-of-freedom challenge when many mutual funds have short sample histories. In contrast, as shown below, our model only requires only three fund-specific parameters; one captures market level timing, one captures volatility timing and one captures selectivity.[8] We must also estimate the market-wide parameters a and b and the mean $E(r_B)$, but these are identified from the benchmarks as described below, and are the same for each fund.

3.1 Relation to Previous Holdings-based Performance Measures

To relate our measure more explicitly to earlier holdings-based measures, use $x'\beta = w'$ and Equation (5) to see that:

$$\mathrm{Cov}(x'r) = \mathrm{Cov}(w'r_B) + \mathrm{Cov}(x'v). \tag{7}$$

This shows that the GT measure has a level timing component, proportional to the first term of Equation (6), and a selectivity component, similar to a part of our selectivity measure. The GT measure leaves out the two terms related to information about second moments. There is no volatility timing term, and the second moment component of selectivity is missing. This is because the GT measure is developed under joint normality with homoskedasticity, so an informed manager never gets a signal that second moments will change. If time-varying

---

decompose the total performance without the portfolio weight.

[8] If a manager times market volatility traditional returns-based measures are even more complicated. Consider a generalization of the Admati et al. (1986) model where the manager gets a signal that is informative about the level of the market and also about its future volatility. The portfolio weight is then related to the future market return and its square, and the portfolio return – the product of the weight and the market return -- is related to the market as a cubic function. This adds cubic factors to the Treynor-Mazuy

second moments are important, our measure offers new and important improvements over the original measure.

Daniel, Grinblatt, Titman and Wermers (DGTW, 1997) develop a popular holdings-based measure where each security *i* held in a fund gets its own benchmark return, R_t^{bi} at each period, t. The benchmark is chosen for each stock from a set of 125 portfolios, as the portfolio that most closely matches the size, book-to-market ratio and lagged return of the stock. In addition, the fund is assigned a set of benchmark weights equal to its actual weights reported k periods before: $x_{i,t-k}$. The DGTW measure is:

$$DGTW_{t+1} = \Sigma_i\, x_{it}\,(R_{i,t+1} - R_{t+1}^{bi}) + \Sigma_i\,(x_{it}\,R_{t+1}^{bi} - x_{i,t-k}R_{t+1}^{bi(t-k)}) + \Sigma_i\, x_{i,t-k}\,R_{t+1}^{bi(t-k)}\,, \qquad (8)$$

where $R_{t+1}^{bi(t-k)}$ is the benchmark return associated with security i at time t-k. The first term is interpreted as "characteristic selectivity (CS)," the second term as "characteristic timing (CT)" and the third as the return attributed to the style exposure (AS). Note that the expected sum of the three terms is equal to the original GT measure, so Equation (8) is a decomposition of the GT measure, and the sum of the terms has the same theoretical justification as the original measure. In particular, it leaves out the same second moment terms that appear in our new measure.

If we take the security specific benchmark, R_{t+1}^{bi} in (8), as an analogue to the systematic component of returns, $\beta\, r_B$ in our Equation (5), then the CS term measuring selectivity in DGTW is analogous to Cov(x'v), the GT selectivity term in Equation (7). The DGTW CS measure leaves out the higher moment part of selectivity. If the second moment effects are

---

regression, further expanding the number of regressors.

important, the inference about fund ability will differ. We find empirically that this difference can lead to different inferences about selectivity performance.

The original GT measure and the DGTW measures use unconditional covariances, and may be misspecified if conditional covariances given public information are important. The evidence in Ferson and Khang (2002) suggests that conditional weight-based measures are important. We therefore extend our measures to consider conditioning information below. We first describe estimation for the unconditional case. Then, the conditional case is a simple extension.

3.2 Estimation

We estimate the market-wide parameters a and b through the short-term Treasury return R_f and the excess return of the benchmarks, as shown in Equations (9a-9b) below. For each fund we estimate a market level timing component, denoted as α_m, a volatility timing component, α_σ, and a selectivity component, α_S. The total alpha for each fund is then $\alpha_p = \alpha_m + \alpha_\sigma + \alpha_S$. The model is estimated using the Generalized Method of Moments (GMM, Hansen, 1982) through the following moment conditions:

$$\varepsilon_1 = (a - b' r_B)\, r_B \tag{9a}$$

$$\varepsilon_2 = (a - b' r_B)\, R_f - 1 \tag{9b}$$

$$\varepsilon_3 = r_B - \mu_B \tag{9c}$$

$$\varepsilon_4 = \alpha_m - a(r_B - \mu_B)' w \tag{9d}$$

$$\varepsilon_5 = \alpha_\sigma + b'(r_B r_B')w - a\, \mu_B' w \tag{9e}$$

$$\varepsilon_6 = \alpha_S - [(a - b' r_B)\, v' x]. \tag{9f}$$

In Equation (9e) we use the condition $0 = a\, E(r_B)' - b\, E(r_B r_B')$ to avoid the need to estimate the parameters of the matrix $E(r_B r_B')$. Because the condition holds exactly at the parameter values that satisfy (9a) and (9b), no additional restrictions are imposed in using this condition. In Equation (9f), we use $v = r - \beta r_B$, where β is the N x K matrix of bottom-up betas estimated using daily data for the stock returns and the benchmarks and v is an N-vector of the idiosyncratic returns of the stocks held.[9] The moment conditions state that $E(\varepsilon)=E(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4,\varepsilon_5,\varepsilon_6)=0$. We use the optimal GMM standard errors (Newey and West (1987) with three lags in the covariance matrix) with the delta method to get standard errors for the sum of the components of performance.

The GMM system (9) is exactly identified and has a block diagonal structure with respect to the fund-specific performance parameters. Results in Farnsworth et al. (2002) imply that the estimates of performance for each fund, when the system is estimated separately for each fund as we do here, are numerically identical to using a full system with many funds stacked together, which is not feasible. If there is public information, Z, we can interpret all of the equations' expectations as conditional on Z. The parameters a and b will also be functions of Z. We discuss such conditional models below.

3. The Data

We study fund data for 1984-2010 from the Center for Research in Security Prices Mutual Fund database. We start the analysis in 1984 because during 1962-1983 there is a

---

[9] Each of the individual betas in $x'\beta$ is estimated by regression using daily data, and the system misses the estimation error in the daily bottom up betas, which we essentially take as data. While betas are estimated with vastly greater precision than alphas, especially

selection bias where funds that report monthly returns have higher average reported returns than the broader universe (e.g. Fama and French, 2010). We exclude fixed income, international, money market, sector and index funds,[10] focusing on active, US equity funds. In some of our analyses we also use daily fund return data from CRSP, available starting in 1998. We subject the fund data to a number of screens to mitigate omission bias (Elton Gruber and Blake 2001) and incubation and back-fill bias (Evans, 2010). In particular we exclude observations prior to the reported year of fund organization, and we exclude funds that do not report a year of organization or which have initial total net assets (TNA) below $15 million in their otherwise first eligible year to enter our sample. We combine multiple share classes for each fund, focusing on the TNA-weighted aggregate share class.

We study US open-ended Equity, Asset Allocation and Balanced funds. These broad groups are determined using the investment objective codes from CRSP.[11] To avoid a possible look-ahead bias due to strategic reporting of investment objectives (Sensoy, 2009) we use the most recent, previously reported code to categorize the funds. When we use holdings data we merge the CRSP and Thompson holdings data using MFLINK and we lose about 8% of the funds (4% of the TNA) due to missing links. To compare with previous studies that focus only on the equity holdings of funds, we report some

---

with daily data, this caveat should be kept in mind when interpreting our empirical results.

[10] We identify and remove index funds both by Lipper objective codes (SP, SPSP) and by searching the funds' names with key word "index."

[11] US equity funds are defined as those with policy codes CS, Flex, I-S; Weisenberger objective codes GCI, IEQ, IFL, LTG, MCG, SCG, G, G-I, G-I-S, G-S, G-S-I, GS, I, I-G, I-G-S, I-S, I-S–G, S, S-G-I, S-I, S-I-G; SI objective codes AGG, GMC, GRI, GRO, ING, SCG; or Lipper objective codes CA, EI, EIEI, ELCC, G, GI, LCCE, LCGE, LCVE, LSE, MC, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, MR, S, SCCE, SCGE, SCVE, SESE, or SG; Asset Allocation funds are identified as funds with Weisenberger objective codes AAL; SI objective codes CPF, EPR, FLX, IMX or Lipper objective code FX. Balanced style funds are identified as

experiments on the equity holdings alone, and some experiments using the active equity funds only. However, given our focus on market factor and volatility timing, we include the asset allocation and balanced funds, and we measure their cash and bond holdings.

Table 1 present summary statistics of the funds' characteristics. The average over time of assets under management ranges across funds from under 300$ million to over 2.4 $billion. The average turnover is almost 90% per year, and is higher for the Asset Allocation style funds. The average expense ratio is 1.1 - 1.25% per year. The funds hold primarily US stocks, although the average Asset Allocation fund holds more than 10% cash and the average balanced fund holds 35% in bonds. For some of our analysis we group funds according to characteristics, including the expense ratio, fund size (TNA), age, turnover, return gap, active share and factor model regression R-squares, described below.

We use daily returns data to estimate betas for the individual stocks held by the funds. The returns data are from CRSP. Our bond index is the Barclays US Aggregate bond index return. This is a value-weighted index of government and investment-grade corporate issues that have more than 1 year remaining until maturity. We obtain daily data for the CRSP stock market factor, the Fama–French (FF, 1996) factors and the UMD momentum factor, from Kenneth French's web site.

5. Empirical Results

5.1 Results for Fund Groups

We first estimate the models on broad groups of mutual funds: Asset Allocation, Balanced and the full active US Equity sample. Since we would not expect to find

---

those with policy code: Bal; Weisenberger objective code BAL; SI objective code BAL or significant performance at the level of broad groups, these exercises serve mainly to check the model specification. We ask if the estimated parameters seem economically reasonable, if the performance estimates look plausible compared with previous research, and if the estimates seem reasonably precise. The results of these exercises are summarized here and reported with tables in the Appendix.

We start with a simple case using the CAPM and a focus on asset allocation, estimating only the equations (9a) – (9e) at a quarterly frequency. We take the asset allocation weights directly from CRSP, which reports the percentage holdings in stocks, bond and cash on an annual or quarterly basis. We use the most recently-available holdings. The estimate of the market-wide parameter a is strongly significant, at 1.03, and the estimate of the parameter b is 2.43 with a t-statistic of 1.82. In the CAPM, the value of a is the inverse of the gross risk-free rate plus a risk adjustment, while the value of b is a version of relative risk aversion, discounted by a pure time preference parameter. These values seem economically reasonable.

The performance estimates suggest insignificant "negative" market level timing ability for the active equity group, and insignificant positive ability for the Asset Allocation group, which is similar to many previous studies using other methods. It makes sense to find more market timing in the asset allocation style funds.

Our second example is a two-factor SDF model with a stock market index and the bond index. The funds' weights are measured as the fractions reported by CRSP for holdings in stocks and in bonds, normalized to sum to one minus the reported holdings in cash. The point estimate of the parameter a is 1.19 and the point estimate of b for the

---

Lipper objective codes B, BT, MTAC, MTAG, MT or AM.

market index is 2.47, both similar to the previous exercise. The value of the *b* parameter for the bond index is 16.76. All of these coefficients are statistically significant.[12]

The performance estimates suggest insignificant "negative" market level timing and positive volatility timing ability, although the economic magnitude is less than 16 basis points per year. The combined timing measure is almost exactly zero for balanced funds and is 7 basis points per quarter or less in each case. The negative level timing results are essentially similar to what previous studies of unconditional timing ability find through other methods. However, negative level timing makes little economic sense and studies find neutral timing for broad fund groups when the model accomodate time-varying risks (Ferson and Schadt, 1996) and other sources of nonlinearity (Chen, Ferson and Peters, 2010). When our models accomodate both level and volatility timing, the combined timing effect is neutral.

We next consider multiple benchmarks and use the full holdings data to measure the portfolio weights in individual stocks. We combine this with the CRSP data and normalize the equity weights to sum to one minus the CRSP reported holdings in cash plus bonds. (Below, we examine results for the equity holdings only.) The asset allocation weights for the benchmark factors are now derived "bottom up" from the individual stock holdings and individual stocks' betas, estimated using daily data over the full available sample for each stock. (Later, we use rolling methods that do not assume constant betas over such long intervals.) We consider two multifactor benchmarks: the FF3 factors and the Carhart (1997) 4-factor model. We also examine a five-factor model including a bond factor. The

[12] In a two-factor Merton (1973) model the b coefficient for bonds depends on the elasticity of the marginal utility of wealth with respect to the bond factor relative to the variance of the bond factor, the latter being a relatively small number which scales up value of the

point estimates of the parameter a and the b coefficient for the market factor are similar to the previous cases, but the b coefficients on the HML and SMB factors are not statistically significant.

The timing performance results for the broad fund groups are essentially similar to the results using only the asset allocation weights. The selectivity term is either negative (for balanced funds) or positive, small and statistically insignificant. The total alphas are economically small. The standard errors say that the alphas are reliably close to zero for the fund groups.

Previous studies find that inferences about performance can be sensitive to the effects of public information. Ferson and Schadt (1996) and Becker et al. (1999) find that market timing ability looks better in models that account for public information. We examine conditional versions of the models, described in the Appendix. We find that moving to a conditional version of the CAPM removes any evidence of negative timing ability, consistent with Ferson and Schadt (1996) and Becker et al. (1999). The standard errors say that timing ability of the fund groups is reliably close to zero in the conditional CAPM.

Previous studies of holdings-based performance often consider only the equity portion of the portfolio. In order to better reflect potential market timing or asset allocation behavior, we include the cash and bond holdings. We conduct some experiments using the conditional models to see if our results are sensitive to this issue. Using the equity holdings only we obtain similar results.

Overall, the analysis at the level of broad fund groups suggests that our models are

---

parameter *b*.

reasonably well specified. The parameter estimates are economically reasonable in magnitude. The performance estimates display patterns similar to those found for broad groups in previous studies using other methods. The estimates also appear to be much more precise than the alphas in returns-based models. We now turn to the evidence at the individual fund level.

5.2 Sorting Funds for Volatility Timing

It may not be surprising to find no significant performance at the level of large groups of funds, but there may be funds with certain characteristics that have performance. In particular, we are interested in the second moment effects that distinguish our measures. In this section we sort funds based on proxies for the likelihood that they actually engage in volatility-related behavior. We estimate these proxies following previous work that uses returns-based methods.

The closest antecedent for volatility timing is Busse (1999). The following regression for funds' daily reported returns generalizes his specification:

$$r_{pt+1} = a_p + \beta_{p0}' r_{Bt+1} + \beta_{p1}' r_{Bt} + [\lambda_{p1}\, \sigma_{mt} + \lambda_{p2}\, (r^2_{mt+1} - \sigma^2_{mt}) + \lambda_{p3}\, (\Delta\sigma_{mt+1})]\; r_{mt+1} + \varepsilon_{pt+1}. \qquad (10)$$

The first two beta vectors in this regression control for the unconditional loadings on the four Carhart (1997) benchmarks. We include the lagged benchmark returns in daily data to allow for nonsynchronous trading. The market beta is modeled as a constant plus the term in square brackets [.]; thus the market beta is linear in several terms including the market volatility, σ_{mt}, which we model as in Busse using an EGARCH (1,1) model on the daily data.

The coefficient λ_{p1} measures the extent to which a fund reacts to market volatility by contemporaneously changing its market beta. This term was also examined by Busse. The coefficient λ_{p2} captures volatility timing in a manner consistent with our model, measuring the extent to which a fund anticipates unexpected changes in the second moments. The final coefficient, λ_{p3}, which was also examined by Busse (1999), is included as a check on the specification, capturing the ability of a fund to anticipate changes in the conditional volatility as captured by the EGARCH model.[13],[14]

Each quarter we select funds with at least 60 daily return observations over the past year and estimate the coefficients of the regression (10). We then sort funds into deciles according to the coefficient estimates and examine the future performance of the portfolios over the next quarter, rolling the whole procedure ahead each quarter. The sample period is 1998-2010 and the results are shown in Table 2.

In Panel A of Table 2 funds are sorted on λ_{p1}, measuring their reaction to market volatility. The differences in the reactions to volatility seem large, as shown in the right hand column. The lowest λ_{p1} funds have negative coefficients, meaning that they reduce market exposure when market volatility is high. Such volatility reaction indicates active management, because a passively managed fund should not have a lower market beta simply because market volatility is high. Previous studies such as Kacperczyk, Sialm and Zheng (2005, 2008), Cremers and Petajisto (2009) and Amihud and Goyenko (2013) find that active management, captured with various measures, is associated with better fund

---

[13] Motivated by nonsynchronous trading we also examine modifications of the regression (10) where a lagged version of the term in square brackets [.], multiplied by r_{mt}, is included and we use the sum of the λ coefficients for the two lags as the sorting parameters. The results are similar to those reported in the table except where noted.

[14] We also replace the daily estimated EGARCH volatility with the VIX index and find

performance.

Table 2 shows that the funds that react more actively to volatility record the best subsequent performance. The total alphas and the selectivity measures after portfolio formation are significantly different across the extreme deciles, with t-ratios of -2.4 and -2.7, respectively. The difference in subsequent total performance between the extreme deciles is 1.1% per quarter. Interestingly, about half of this is due to the selectivity component, and half to the combined timing ability.[15] Thus, funds' reaction to market volatility can detect active management associated with stronger performance.

The standard errors in Table 2 confirm our previous impression of good precision for both the selectivity component and the total alpha. The standard errors for these measures are usually between 1-2 basis points per quarter for each decile group. The precision of the timing estimates does not seem to be as high. The differences in timing ability associated with volatility reaction shows a spread of 49 basis points per quarter across the extreme deciles, with the funds that reduce exposure in response to market volatility scoring larger market level timing coefficients. This difference, however, is not statistically significant. We also examine the DGTWcs measures for these portfolios, and the lower λ_{p1} funds have the larger selectivity measures, but the differences are not statistically significant. Thus, our measures appear to ferret out ability that the DGTWcs measure does not detect, when funds are sorted according to their proclivity for reacting to time-varying market volatility.

In Panel B of Table 2 we repeat the exercise sorting funds on their λ_{p2} coefficients,

---

essentially similar results.

[15] When we include the lagged interaction terms in the regression (10) the selectivity effect is 70 basis points with a t-ratio of 2.1, but the total alpha difference is no longer significant.

which measure the ability to anticipate unexpected volatility. A negative coefficient describes a fund that gets out of the market before volatility rises. In particular, the fund anticipates the squared market residuals better than the information in the EGARCH model. In this case, the differences across the portfolios in their subsequent performance measures are not significant.

It might be too much to expect a mutual fund's holdings to predict market volatility better than a daily EGARCH model. In Panel C of Table 2 we sort on the λ_{p3} coefficients, which capture the alternative measure of timing ability based on changes in the EGARCH conditional volatility. The funds with stronger volatility timing record significantly better subsequent selectivity performance. The difference between the selectivity measures for the extreme deciles is about 2.8% per year and the t-ratio is larger than three.[16] The total performance differences are not significant and the DGTWcs measures present no statistically significant differences.

The results in Table 2 provide evidence that funds whose reported returns suggest a greater proclivity to trade actively in response to volatility have better subsequent performance for the next quarter. Selectivity is the dominant component of the performance, but the DGTWcs selectivity measure shows no significant relation to funds' volatility-related behavior. We investigate the selectivity further by looking separately at its two parts. For the evidence in Panel A of Table 2 we find that neither of the two terms displays a significant spread by itself, while the combined effect is significant. This

---

[16] When the lagged interaction terms are included in the regression (10) the difference is about 2% per year. We also estimate versions of regression (10) where we constrain some of the λ-coefficients to be zero and estimate only the coefficients of interest. In another version we use the funds' equity holdings only, assuming these weights sum to 1.0. The results of these experiments are broadly similar.

suggests that the DGTWcs measure finds no relation because it leaves out the higher moment effect. For the evidence in Panel B there is no significant relation, while in Panel C the selectivity performance is mainly driven by the first moment term.

5.3 Sorting Funds on Factor Model R-squares

Amihud and Goyenko (2013) find that when funds are sorted according to the regression R-squares of their returns in standard factor models, the funds with lower R-squares have higher subsequent performance. Titman and Tiu (2011) find that low R-square hedge fund perform better. Low R-squares likely indicate more active management; either departing from a benchmark or focusing on stocks with larger fractions of variance attributed to firm-specific information.

Table 3 sorts our sample of funds as in Amihud and Goyenko, using daily reported fund returns to estimate the factor model R-squares. Consistent with Amihud and Goyenko, the DGTWcs measure is larger for the lower R-square funds, by about 24 basis points per quarter, but this is not statistically significant. Our selectivity measure is also higher for the lower R-square funds, of a similar magnitude and not statistically significant. Our overall alpha measures vary substantially with the R-squares and the difference sports a t-ratio over two. This difference between the extreme deciles, about 3.2% per year, is likely economically significant. The overall performance spread reflects selectivity (about 62%) and combined timing ability (about 38%). Thus, our new measure, by accomodating the response to changes in second moments that is likely of more active funds, produces sharper estimates of how performance varies with fund activity.

One possible concern with Table 3 is the role of the balanced and asset allocation funds. If the risks of such funds are poorly captured by the Carhart (1997) model, they may have both low R-squares and overstated alphas. We replicate the exercise in Table 3, removing the balanced and asset allocation funds from the sample. The DGTWcs measure is larger for the lower R-square funds, by about 30 basis points per quarter, but not statistically significant. Our selectivity measure is higher for the lower R-square funds. The spread is 70 basis points per quarter, but not statistically significant. Our overall alpha measures differ by 3.6% per year across the deciles, with a t-ratio of 1.87, so the results are very similar to the results that include the asset allocation and balanced funds. We also conduct an experiment where we use a five-factor model, including the Barclays bond index factor, to compute the R-squares and the performance for the full sample of funds. The spread in total alphas across the R-square deciles is 3.2% per year with a t-ratio of 2.11. The spread in the DGTWcs measure is only 1.1% per year, with a t-ratio of 1.48. Thus, the findings that the low R-square funds have better total performance, and that this not detected by the DGTWcs measure, are not driven by the asset allocation and balanced funds nor by the choice of factor models.

Amihud and Goyenko (2013) and our Table 3 use the after-cost daily reported fund returns from CRSP to compute the factor model R-squares. We check the sensitivity of our results to an alternative measure of R-squared, based on the hypothetical returns formed from the portfolio holdings and the daily returns of the underlying stocks. The sample period for this exercise is 1995-2010. In results not tabulated here, we find that the results are similar. The differences between the two versions of factor model R-squares could be affected by funds' actions between the portfolio holdings' reporting dates. One R-square measure uses the daily returns, which reflect the actions between the quarterly holdings dates, and the other R-square uses the hypothetical returns, based on the quarterly holdings. But the two samples of funds do not perfectly overlap. To control for that difference, we sort on the two measures of R-squared, restricting to the subset of funds for which both types of R-squares can be computed. The differences between our total alphas for the extreme deciles sorted on R-square are 80-90 basis points per quarter and the corresponding absolute t-ratios are slightly larger than two, using either version of the R-squared. So, the Amihud and Goyenko (2013) finding seems robust to the measure of R-squared, at least in the subsample where both versions can be calculated.[17] Overall, the results in tables 2 and 3 show that our new measures detect ability associated with active management in cases where the DGTWcs measure cannot.

5.4 Ex-Post Abnormal Performance

Recent papers such as Kacperczyk, van Nieuwerburg and Veldcamp (2012) examine *ex post* measures of performance. For example, in a linear factor model regression the intercept plus residuals define an abnormal performance measure at each date. This approach may have more power to detect abnormal returns, because there may be patterns in the residuals that average to zero and are ignored by the intercept. We examine a version of this approach. We use the most recently reported holdings and a buy-and-hold assumption to build a monthly series of portfolio holdings, known at the end of each month, t. These are applied to returns for month t+1. First, we estimate the benchmark parameters (a,b,μ_B) like before using the subset of moment conditions (9a-9c). We use the

---

[17] Sorting on the difference between the two versions of R-squared for each fund, we find no significant performance differences across the deciles. This further suggests that the relation between performance and the R-squares from daily reported returns is not driven

previous 36 months' data on the benchmarks for this estimation. We use the past year of daily data to estimate the factor betas of the underlying securities held, and we compute a bottom up asset allocation weight, w_t, for each month. Then, we compute three monthly performance series for each fund for the subsequent month:

$$LT_{pt+1} = a(r_{Bt+1} - \mu_B)' w_t, \tag{11a}$$

$$VT_{pt+1} = [- b'(r_{Bt+1} r_{Bt+1}') - a\, \mu_B']w_t, \tag{11b}$$

$$SEL_{pt+1} = [(a-b' r_{Bt+1})\, v_{t+1}' x_t]. \tag{11c}$$

We use the ex post versions of our performance measures for several experiments.

5.5. Predicting After-Cost Fund Returns

While holdings-based measures are useful for evaluating manager skills before costs, it is interesting to see if the measures contain information about the returns available for investors after fund costs. According to the model of Berk and Green (2004), differences in before-cost ability should be offset by differences in costs across funds, resulting in no abnormal performance for investors.

Each quarter we sort funds into quintiles based on our ex post measures of alpha, using data up to the end of the quarter. We compute five equally-weighted portfolios based on the quintiles and examine their reported returns for the next three months. Rolling this procedure forward through time, we generate three time series of returns for each quintile; one for each of the three months after portfolio formation. We find that the excess reported returns are higher for the high-performance quintiles in each subsequent

---

by funds' actions between reporting dates.

month. The differences for the extreme quintiles are as large as 15 basis points per month. The differences in the Carhart (1997) four-factor alphas for the after-cost returns are as large as 16 basis points per month. None of these differences is statistically significant, reflecting the large standard errors of the returns-based performance measures. However, using our total performance measure as the sorting criteria, the spread portfolios' returns and alphas are monotonic across the quintiles, while the sorting on the DGTWcs measure we find random patterns in the subsequent performance.

Possibly, the changes in funds' portfolio weights induced by trades could be more informative about future after-cost returns than are the holdings levels (e.g., Chen, Jegadeesh and Wermers, 2000). We repeat the previous analysis, replacing the levels of the portfolio weight for each quarter with the difference between the reported weight and what it would have been if the fund had used a buy-and-hold strategy since the previously-reported holding. Similar to the previous analysis, the higher-performance quintiles have higher subsequent returns in most cases, but the differences are not statistically significant. The differences in the extreme quintile returns are 13 basis points per month or less and the Carhart four-factor alphas are 12 basis points or less. Sorting on the DGTWcs measure, insignificant random patterns in the future returns are observed across the quintiles. Overall, our holdings-based measures are better at selecting funds with subsequently higher after-cost returns than is the DGTWcs measure, but neither measure leads to a significant rejection of the Berk and Green (2004) prediction of zero after-cost alphas.

5.6 Conditional Analyses of the Ex Post Performance Measures

The *ex post* performance measures from system (11) may be useful for detecting performance conditional on different economic states, because it includes performance fluctuations that are averaged out over time in the moment conditions of system (9). By averaging the residuals separately within the given states, we could gain power to detect differences across the states.

We first run regressions of the ex post abnormal performance measures on the state of market sentiment, as measured by the Baker and Wurgler (2006) sentiment index at the end of the previous month. The results are summarized in Table 4. In Panel A funds are sorted into quintiles based on their factor model R-squares over the past 36 months. All of the slope coefficients for the level timing ability measures and for the total alphas are negative, sport absolute t-ratios larger than two in eight of the ten examples and the t-ratios are greater than 3.0 in three of the ten cases. This suggests that funds are better (less) able to time the market when investor sentiment is low (high). The effect is weaker for the more active, low R-square funds, but still statistically significant. The average coefficient, -0.022, suggests that the effect of timing ability on alpha varies by about 1.3% per month in association with a one-standard deviation move in investor sentiment. The fitted values of the expected conditional alphas change sign with sentiment. They are all negative when sentiment is one standard deviation above its mean, and all positive when sentiment is one standard deviation below its mean. The level timing effect of sentiment dominates and largely determines the coefficients of the total alphas on the sentiment index.

In Panel B of Table 4 we sort funds on the volatility reaction coefficients, γ_{1p}, from Equation (10) estimated over the past 36 months. We find that all of the slope coefficients for total performance and level timing ability are negative, with t-ratios larger than two. The coefficient magnitudes are similar to Panel A, and similar across the quintile groups.

Novy-Marx (2012) raises concerns about spurious regressions for stock returns using the sentiment index. We provide empirical p-values for the t-ratios in Table 4 using a version of the parametric bootstrap. We model the vector $\{\alpha_t, Z_t\}$ as a VAR(1) process, where Z_{t-1} is the lagged sentiment index and α_t is the ex post performance measure, and take the coefficients as parameters. We resample the residuals from this model at random, without replacement, and add them sequentially to the fitted values of the conditional means from the VAR, building up the simulated time-series recursively. We impose the null hypothesis that the slope coefficient for α_t on Z_{t-1} is zero and set the intercept for α_t equal to its unconditional sample mean. This preserves the high autocorrelation of the sentiment index, equal to 0.96 in our monthly sample. The empirical p-value is the fraction of 1,000 simulation trials in which the t-ratio using the simulated data is larger in absolute value than in the original sample (a two-tailed test). The empirical p-values confirm that, while there is a modest size distortion in the regressions, all of the slope coefficients for ex post timing ability on the sentiment index are significant at the 5% level or better.

One potential interpretation of this evidence that market timing ability differs conditional on sentiment is consistent with the cash flow mechanism described by Ferson and Warther (1996). Funds receive more new money flows when market sentiment is high (Ferson and Kim, 2012). Ferson and Warther find that aggregate new money flows can help explain “negative timing.” When new money flows are large it may be more costly to try to time the markets. When sentiment is low and money flows are muted, it may be less costly for funds to engage in market timing activity. A similar argument is implicit in Edelen (1999), where fund performance on trades in response to flows is worse than it is on “discretionary” trades that are not made in response to flows.

To see if new money flows explain the sentiment effect on performance, we run regressions of the ex post performance on the lagged sentiment index and a lagged value-weighted aggregate of the new money flows for the funds in our sample. The new money flows enter the regressions with negative coefficients, and the empirical p-values are below 10% in nine of the ten examples, but the flows do not subsume the explanatory power of sentiment. The regression slope for timing ability on sentiment remains negative and is significant at the 5% level in every case. These regressions suggest that the cash flow mechanism of Ferson and Warther (1996) and Edelen (1999) may be in effect, but it does not fully explain the role of sentiment. Whatever the mechanism, the regressions show that the ex post performance measures vary significantly across economic states, as captured by the sentiment index and the aggregate new money flows.

We examine additional market states, including the level of idiosyncratic volatility, recessions versus expansions and levels of market volatility. The average level of idiosyncratic volatility may indicate times when the environment favors firm-specific information. When common factors are calm and firm returns are driven to a greater extent by idiosyncratic factors, there is likely to be more room to pick stocks. We measure the (complement of the) average idiosyncratic volatility by the average of the R-squares of individual stocks in daily Carhart (1997) factor model regressions. The average of the firm R-squares is updated at the end of every quarter.[18] Regressing each component of the broad fund groups' *ex post* abnormal returns over time on the predetermined average R-squares, none of the 15 regression slope coefficients delivers a t-ratio larger than 2.0.

---

[18] At the end of each quarter we screen out stocks with extreme returns (>200%) or infrequent trading (zero returns on ≥ 75% of the days) during the quarter. We run the regressions for the remaining stocks over the days of the quarter and average the regression

However, all nine of the timing measures have positive coefficients and all three of the selectivity coefficients are negative, so there is a hint that it is harder to time and easier to pick stocks when the average firm's idiosyncratic volatility is relatively high.

Perhaps, subsets of funds and more active funds in particular, are more likely to display an idiosyncratic volatility effect. We there examine performance conditioned on the average idiosyncratic volatility measure for funds grouped into quintiles on the R-squares in daily factor model regressions for the reported returns, on the Busse (1999) volatility reaction coefficients, γ_{p1}, described above, and on the volatility timing coefficients, γ_{p2}, from Equation (10). With five quintile portfolios for each of five performance measures, there are 25 t-ratios for each of the three sorting criterion, and a total of 75 performance test statistics. Four of the absolute t-statistics are greater than two and the maximum absolute value is a t-statistic of -2.657. All of the coefficients, except for the level timing effect, are negative. So again there is a hint that selectivity performance is better when firm-level idiosyncratic volatility is high, but with so many tests the overall evidence is not strong.

When multiple tests are examined, the Bonferroni p-value is a conservative bound on the two-tailed p-value for the hypothesis that none of the performance measures is different from zero. This computed by multiplying the p-value of the maximum absolute t-ratio by the number of tests.[19] In the previous example the Bonferroni p-value is 59%. Thus, based on the Bonferroni p-value the evidence for the effect idiosyncratic volatility effect on ability is not statistically significant.

Studies such as Moskowitz (2002), Ferson and Qian (2004), Kowsowski (2011),

---

R-squares across the stocks.

[19] If A and B are rejection events for two statistics, P(A or B) = P(A) + P(B) - P(A and B). The Bonferroni bound says P(A or B) ≤ P(A) + P(B) and uses the smallest p-value multiplied by

Glode (2011) and Kacperczyk, van Nieuwerburg and Veldcamp (2012) suggest that fund performance may vary over the state of the business cycle, with stronger performance in recession periods. We use the *ex post* abnormal fund performance measures to examine performance over the business cycle, defined by NBER reference dates. At the broad fund group level, breaking the sample into recession and expansion subperiods, there are 30 tests for performance and none of the absolute t-ratios are larger than 2.0. When we group funds into quintiles by the three criteria described above and regress each portfolio's abnormal performance on a recession indicator dummy, we find no significant performance numbers.

If the before-cost ability does not vary across business cycles, while after-cost performance does according to the previous studies, it suggests that there is cyclicality in the cost structures of mutual funds. Teasing out these patterns seems to be a topic worth future research.

We also condition performance on market volatility states, as measured by the VXO volatility index. Regressing the three broad fund groups and the three sets of quintile groups on the VXO at the end of the previous month, there are 90 performance tests. The largest absolute t-ratio is a t-statistic of 1.84, providing no evidence that performance varies across the VXO states.

5.7. Sorting Funds on Other Characteristics

Previous studies have identified other fund characteristics associated with performance, and we consider several such characteristics in this section. These include the expense ratio, turnover, active share and return gap, defined below. Many of these

---

the number of tests to bound the joint probability.

characteristics are likely to be associated with the differences between funds' reported returns and the holdings-based returns that are the focus of our measures, such as costs and interim trading effects. Many of the previous studies that emphasized these sorting criteria used the "characteristic selectivity" measure from Daniel, Grinblatt, Titman and Wermers (1997). We also report results for this measure, DGTWcs in the tables, and note how the results differ from the results for our measures. We use the Carhart 4-factor benchmark model for these exercises. This controls for size, book-to-market and momentum factors, corresponding to the characteristics used in the DGTWcs measure.

One interesting comparison between the performance measures is their correlations across individual funds. Traditional performance measures tend to be highly correlated across funds (e.g. Lehman and Modest, 1987), but our measures are not highly correlated with the DGTW measures. Depending on the screens applied to the fund samples, we find correlations between our selectivity measure and the DGTWcs measure of 0.19-0.45. Our level timing measure has correlations with the DGTW characteristics timing (CT) measure of 0.12-0.31, and our volatility timing measure has correlations that range between -0.10 and -0.21. While low correlations can reflect estimation noise, the preceding analysis shows that our measure detect performance and predict returns in cases where the DGTW measure does not. This suggests that the additional second moment terms that our measures bring to the analysis differ across funds from the previous measures.

In panel A of Table 5 we sort funds into equally-weighted deciles according to their most recently-reported expense ratios. The DGTWcs measures are slightly higher for funds with higher expense ratios, as previous work has shown, but this pattern is not statistically significant. Using our measures, the selectivity performance is actually lower for the high-expense-ratio funds, by about 8 basis points per year, with the difference again not statistically significant. The estimates suggest that the higher moment terms in our selectivity measure, ignored by the DGTWcs measure, vary by about 10 basis points per year across the expense ratio-sorted deciles. Our total performance measure is also higher for the low expense ratio funds, by about 4 basis points per year, and not statistically significant. The difference between the total performance and the selectivity is the market timing terms, which differ by about 4 basis points across the deciles.

Cremers and Petajisto (2009) propose an “active share” measure, the mean absolute difference between the holdings of a fund and the holdings of the benchmark. Sorting funds on this measure, they find that excess after cost reported returns differ significantly, by about 2.5-3% per year across quintiles, and the more active funds deliver higher future returns net of a fund-specific benchmark. Petajisto graciously provided data for the active shares on his web page. Panel B of Table 5 sorts funds by the active shares and examines our performance measures along with the DGTWcs measure. The DGTWcs measures increase with the active shares, but the differences between the extreme deciles, about 6 basis points per year, is not statistically significant. Our selectivity measures vary in the opposite direction across the deciles, the lower active share funds recording greater selectivity by about 4 basis points per year, not statistically significant. Similar to the results in Panel A, the higher moment component of the selectivity term that is ignored by the DGTWcs measure but included in measure, amounts to about a 10 basis point per year difference across the deciles. The total performance measures are similar to our selectivity measures in Panel C because the level and volatility timing effects roughly offset in the decile differences.

Kacperczyk, Sialm and Zheng (2008) examine direct measures of the effects of portfolio actions between reporting dates. They find that sorting funds by their lagged one-year "return gap," defined as the difference between the reported, after cost return and the hypothetical holdings-based return, can predict subsequent performance using several performance measures. Sialm provides data for funds' return gap on the web page of the Review of Financial Studies for 1984-2006. Sorting by return gap in Panel C of Table 5 for 1984-2006, we find that the DGTWcs measures are larger for the higher return gap funds, by about 4 basis points per year, but the difference is not statistically significant. Our selectivity measure again varies in the other direction across the deciles, but the difference of -4 basis points per year is not significant. The total performance alpha is higher for the high-gap funds, by about 1.4% per year. This is completely driven by the highest gap decile, with more than 2/3 of the effect attributed to level and volatility timing performance. None of these differences, however, is statistically significant.

The main focus of Kacperczyk, Sialm and Zheng (2008) is the after-cost reported fund returns. Using these, we confirm that sorting on return gap delivers a significant spread in the Carhart 4-factor alphas over their sample period. Our measures, without interim trading effects or any costs, likely produces different results than the after cost returns, because interim trading and expenses have a direct impact on after-cost returns and on the return gap.

Wermers (2000) finds that funds with higher lagged turnover generate higher DGTWcs measures in quintile sorts. Panel E of Table 5 sorts funds into quintiles on their most recently reported turnover. The DGTWcs measures are larger for the higher turnover funds, but the pattern is not monotonic or statistically significant. The combined timing measure is also larger for the high turnover funds, but the difference does not attain statistical significance.

As previously mentioned, earlier studies of holdings-based performance often consider only the equity portion of the portfolio. In order to better reflect potential market timing and asset allocation behavior, we include the cash and bond holdings. Possibly, differences in the results compared with DGTWcs, is related to our broader view of the funds' portfolios. Panel E of Table 5 strongly suggests this is not the case. We replicate panel D using the equity holdings only, assuming the equity weights sum to 1.0 each quarter. The results are almost identical. The DGTWcs measures and our alpha measures differ across the Panels D and E only in their last digit.

Cremers and Petajisto (2009) find a negative trend in funds' active shares over time, and suggest that recent data may be influenced by more "closet indexing" among active mutual funds. Kim (2012) finds that the flow-performance relation in mutual funds attenuates after the year 2000, which could be related to volatility that renders recent performance less informative, or to a trend toward more similar performance in the universe of managers. The earlier studies by Grinblatt and Titman (1993), Wermers (2000) and others use data that do not cover the period after 2000. Perhaps, the evidence for investment ability has changed in the more recent data. We repeat the analyses from Table 5, using the restricted 1984-1999 sample period. The results are similar to Table 5. We find no evidence that the different findings using our measures are driven by the post 2000 data.

5.8 A Multiple-comparisons Perspective

The preceding tables have presented many estimates of fund performance. It is important to interpret this evidence in view of the number of tests conducted. We use Bonferroni p-values to evaluate the extreme statistics in some of the discussion. This is a conservative bound, as it sets the joint probability of two rejection events to zero. A simple alternative approach is to evaluate the fractions of the tests that reject the null hypothesis of zero performance at a given significance level. In the tables to this paper we present 318 estimates of performance and obtain 21 absolute t-ratios larger than 2.0. Using simple binomial probabilities and allowing for a conservative amount of correlation across the tests, the overall t-ratio for observing 21 rejections at the 5% level, when the expected number of rejections is 15.9, is approximately 0.04.[20] If we add the estimates discussed in the text but not tabulated, there are 2155 estimates of performance and 124 absolute t-ratios larger than two. The binomial t-ratio for this is -0.002.

We select the preceding tables considering the flow of the paper and to highlight "interesting" results. In the analyses conducted for this paper, our best estimate is that we generated 9054 estimates of performance and found 381 absolute t-ratios larger than 2.0. The binomial t-ratio for finding 381 rejections of the null that performance is zero, using a 5% test, when 9054 tests are conducted is close to zero.

While a multiple comparisons perspective suggests that overall, there is little evidence of ability, we do observe some patterns that suggest some of the results are not

---

[20] If y_i is an indicator variable for the i-th test rejecting at the 5% level and there are N tests with correlation ρ, then Σy_i has mean 0.05N and variance $N(.05)(.95)\{1+(N-1)\rho\}$. The t-ratio in the example is $(21-15.9)/[318(.95)(.05)\{1+317\rho\}]^{.5}$ The result is very sensitive to the assumed value of the correlation, ρ. Using the ex post abnormal performance components described above for each fund, the sample correlations of two components taken across funds ranges from -0.23 to +0.19 and the average is 0.028. At the extreme, if funds are sorted into quintiles randomly, the correlations of the measures across the quintiles are much higher, upwards of 90%. We use ρ=0.03 in these calculations. The t-ratios are smaller if ρ is larger.

random statistical flukes. We should be cautious, as Richardson (1993) and others emphasize that correlated tests are likely to produce the appearance of patterns. One of the marginal cases is when we condition the ex post abnormal performance on the average idiosyncratic volatility. We find that all 18 of the regression coefficients for level timing are positive, 16 of the coefficients for volatility timing are positive and 17 of the selectivity coefficients are negative. If the sampling distribution is symmetric under the null of zero coefficients, we should find half of the estimates are negative and half are positive. Under the same correlation assumption used above, the binomial t-ratio for the number of positive coefficients is 3.45 for level timing, 2.68 for volatility timing and -3.07 for the selectivity coefficients.[21] However, if the fund grouping is completely random the correlation of the estimates would be much higher than assumed in this calculation, as we would effectively have little more information than a single regression, not 18 separate tests. If the correlation is 0.9 all of the binomial t-ratios are close to zero. In this case the average value across the 18 regressions might be a better statistic. The average coefficient has the indicated sign, but is not statistically significant. Thus, there is only weak evidence of stronger timing ability when the average idiosyncratic volatility is low or stronger selectivity performance when it is high.

5.9 Precision and Power

As discussed previously, holdings-based measures could suffer a loss of power for various reasons. Also as mentioned above, Ferson and Khang (2002) and Jiang et al. (2007) examine the power of weight-based approaches with simulation and find that using the

---

[21] For example, using a 50% probability of a positive coefficient, the t-ratio for finding all 18

information in the many portfolio weights offsets the loss of information in a single time series of the reported fund returns, so that holdings-based measures are quite powerful. A full simulation analysis of power is far beyond the scope of this study, complicated by the need to model the portfolio weights of funds with varying degrees of ability and other issues. See Wang (2013) for a discussion and analysis of the statistical properties of holdings-based measures.

It seems that the precision of our estimates is high. For example, in the Appendix tables A.1-A.2 the average standard error of the alpha estimates at the broad fund group level is about 5 basis points per quarter, so performance of 2% per year would earn a t-ratio of about ten. In Table 5 where funds are sorted into decile portfolios, the average standard error of a performance estimate is about 19 basis points per quarter, so performance of 2% per year would earn a t-ratio of about 2.6. In the quintile sorts the standard deviation is about 14 basis points per quarter. This is much more precision than available with returns-based alphas.

While a comparison of our performance estimates with returns-based estimates suggests that our approach delivers good precision, a comparison of our measures with the DGTWcs holdings-based measure is a slightly different story. Averaged across all the panels in Table 5, our selectivity measure's standard error is 16 basis points per quarter and our total performance alpha's standard error averages 21 basis points per quarter. By comparison, the DGTWcs measure's average standard error is 11 basis points. Thus, bringing the timing components and second moment terms in with our measures does present a modest cost in precision.

---

coefficients positive is $(18-9)/[18(.5)(.5)\{1+17\rho\}]^{.5}$

Why should we estimate selectivity simultaneously with the timing and second moment effects at a cost of lower precision? As in many problems, we face a tradeoff between bias and efficiency. If we estimate the full model with all the components, and some of them are really zero, we sacrifice precision on the other components. If we leave out some of the components of performance that are not zero, we face a left-out variables bias for the components we include. The standard errors suggest that our loss of precision, compared with the DGTWcs measure, is modest. Importantly, our evidence shows that the components missing from the DGTWcs measure are not zero. For example, we find that the relation between total performance and volatility reaction in Table 2 is almost 50% driven by the timing effects, and by leaving out those terms the DGTWcs measure finds no significant performance. The relation of performance to the Baker and Wurgler (2006) sentiment measure that we document above is almost entirely driven by the market timing effects, and is not found by the DGTWcs measure. Thus, there is useful information in the components of performance.

5.10. Additional Tests

The preceding evidence suggests that active funds, with activity measured in various ways, have stronger future performance. Another measure of activity is "style drift," in which funds change their style exposures over time. We measure the style drift of a fund using its daily reported returns over the past year to estimate Carhart (1997) four-factor betas for the most recent and the lagged six-month periods, β_1 and β_2. We use the full year to estimate the covariance matrix of the beta estimates, $V(\beta)$.

Our measure of style drift is similar to a Chow (1960) test: $(\beta_1 - \beta_2)'V(\beta)^{-1}(\beta_1 - \beta_2)$.

We sort funds into deciles on the style drift measure and estimate the performance of the decile portfolios over the next quarter, rolling the whole procedure forward quarterly. We find that the more active funds record subsequent higher selectivity, by about 1% per year compared with the least active. The t-ratio of the difference is 2.04 with our selectivity measure and 2.77 using DGTWcs. These tests confirm with yet another measure of activity, that more active funds display better ability. The timing measures suggest that the style drifting funds tend to anticipate high return and high volatility factors, but this is not a significant effect. The differences in the level and volatility timing terms across the deciles is 1.6% and -1.2% per year respectively, so the total timing effect which is the sum of the two terms, is small.

Grinblatt and Titman (1993) and DGTW (1997) study the persistence in their measures. We sort funds on the basis of previous estimates of each of our three components of performance each quarter. These estimates use the previous two years of data. This is a short sample for estimation, resulting in noisy estimates, but requiring that a fund survive for a longer period increases the survival selection bias and can create spurious persistence (e.g. Brown et al. 1992). We find no evidence of persistence in our measures.

De Souza and Lynch (2012) criticize previous studies that find fund performance varies over the state of the business cycle, for using NBER reference cycles like we do above, because these are only known *ex post*. They find that the evidence for business cycle variation in performance weakens substantially or disappears when *ex ante* conditioning variables are used. We find a similar pattern with our conditional measures. We estimate a probit model for the likelihood of a recession and break the sample up into high, medium and low ex ante recession probability subsamples. Any evidence for differences in performance across the business cycle states is even weaker in this exercise.

6. Conclusions

A holdings-based performance measure should reflect the ability of funds' portfolio holdings to anticipate the subsequent abnormal, or risk adjusted returns of the securities held. Generalizing previous holdings-based measures to make the risk adjustment shows that investment performance has components related to factor level timing, volatility timing and selectivity. We develop and implement simple measures of performance that account for all three components, without making any stylized assumptions about manager behavior. Our approach can be used with any specification for the stochastic discount factor, and we illustrate it using popular linear factor models.

Allowing for market level and volatility timing, we find that funds with more active responses to volatility have better investment ability. We also find that the ability to time market factor levels is weaker when an investor sentiment measure is high, and stronger when it is low. This is consistent with a deleterious impact of exogenous fund inflows on performance, but is not fully explained by aggregate fund flows. Sorting funds by factor model R-squares confirms the findings of Amihud and Goyenko (2013) that the low R-square funds have better ability. Measuring activity in the form of "style drift," produces similar results.

Comparing our new measures with popular holdings-based measures of selectivity from Daniel, Grinblatt, Titman and Wermers (1997), we find that the second moment and market timing effects that we include reveal performance, related to active management, that remains hidden from the earlier measures. Our new measures should be especially useful in settings where changes in conditional second moments are an important feature of the investment environment.

Appendix

A.1 Tables of Group Level Estimates

Table A.1

Components of Performance in a Market Timing Setting

This table summarizes results when the unconditional CAPM defines the benchmark, using CRSP data on mutual funds' weights in stock. The sample covers January of 1984 through December of 2010. The GMM with a Newey-West lag of three is used for estimation. Level timing is the estimate of α_m, volatility timing is the estimate of α_σ and their sum represents the total market timing performance. NObs is the number of time series observations of the fund group used for the estimation. Panel A reports GMM estimates of a and b in the system (9).

Panel A:
Estimates of the Market-wide Parameters:

		NObs	a	b	
	Est	108	1.033	2.43	
	t_stat		23.73	1.82	

Panel B:
Performance

			Level Timing	Volatility Timing	Sum
AssetAllocation	**Est**	33	0.0004	0.0002	0.0006
	t_stat		1.2743	1.2986	1.8123
Balanced	**Est**	36	-2E-04	-1E-04	-0.0002
	t_stat		-0.353	-0.27	-0.5473
USEquity	**Est**	55	-7E-04	0.0001	-0.0005
	t_stat		-1.456	0.3683	-1.0373

Table A.2

Components of Performance in an Asset Allocation Setting

This table summarizes results for asset allocation performance in a two-factor setting, where the stock market index and a bond index are the benchmarks, using CRSP data on mutual funds' asset allocation weights in stocks and bonds. The bond index is the Barclays US Aggregate bond index and the stock market index is the CRSP value-weighted market index. The sample covers January of 1984 through December of 2010. The GMM with a Newey-West lag of three is used for estimation. Level timing is the estimate of α_m, volatility timing is the estimate of α_σ and their sum represents the total asset allocation performance. Nobs is the number of time series observations of the fund group used for the estimation. Panel A reports GMM estimates of the parameters a and b in the system (9).

Panel A:
Estimates of the Market-wide Parameters:

		NObs	a	b1(mkt)	b2(bond)
	Est	108	1.19	2.47	16.76
	t_stat		14.26	2.17	3.82

Panel B:
Performance

			Level Timing	Volatility Timing	Sum
AssetAllocation	**Est**	33	-0.0007	0.0008	0.0001
	t_stat		-1.236	1.6648	0.1159
Balanced	**Est**	36	-0.0004	0.0004	0
	t_stat		-0.79	2.0189	-0.008
USEquity	**Est**	55	-0.002	0.0009	-0.0007
	t_stat		-1.583	1.3533	-0.954

A.2 Conditional Models

The conditional models follow Cochrane (1996), assuming that the parameters a and b are linear functions of lagged instruments, Z, and that the conditional means of the benchmark excess returns are linear functions of Z.[22] Thus, a and b are replaced by linear functions a(Z)=a′Z and b(Z)=b′Z, and μ_B is replaced by a linear function $\delta_B Z$ in Equations (9c-9e), where δ_B is a K x L matrix of parameters, a and b are L-vectors of parameters and L is the number of lagged instruments in Z, which includes a constant. The modified equations (9a-9c) are multiplied by each element of Z. The GMM with a Newey-West (1987) covariance matrix using three lags is used in the estimation of the standard errors.

In the conditional CAPM the estimates of the market-wide parameters suggest that the coefficient b(Z) is a time-varying function of the lagged instruments; indicating a time-varying price of market risk, but we do not reject the hypothesis that a(Z) is a constant function over time. The conditional market timing model indicates insignificant overall market timing ability and both components of timing ability are insignificant.

We also examine a conditional asset allocation model with two factors: the market index and the bond index, similar to Table A.2. There is less evidence of time-varying SDF coefficients in this model, consistent with the less-significant b coefficients in Table A.2. The overall flavor of the results is similar to that of the market timing example. We find no significant negative level timing in the conditional version of the asset allocation model. The sum of the two components of timing ability is insignificant and numerically close to zero. We estimate a conditional version of the FF3 factor model and examine all three

[22] We also examine parametric conditional models that assume linear functional forms for the first and second conditional moments of the benchmark returns and derive nonlinear functions for time varying a_t and b_t coefficients from the restrictions of the model. These

components of performance using the Thompson holdings data. The results are broadly similar. We also estimate a conditional version of the Carhart 4-factor model and find essentially similar results.

In summary, the conditional models confirm the evidence for the broad fund groups. Allowing for level and volatility timing behavior, no strong evidence for performance is found at the fund group level. The economic magnitudes of the performance estimates are small, and the standard errors say the performance is reliably close to zero.

References:

Admati AR, Bhattacharya S, Pfleiderer P, Ross SA. 1986. On Timing and Selectivity. *Journal of Finance* 41: 715-30.

Aragon G, Ferson W. 2008. Portfolio Performance Evaluation. *Foundations and Trends in Finance* 2:2: 1-111

Aragon, George O., and J. Spencer Martin, 2012. A unique view of hedge fund derivatives usage: Safeguard or speculation? *Journal of Financial Economics* (forthcoming).

Amihud, Y. and R. Goyenko, 2011, Mutual fund R^2 as a predictor of performance, *Review of Financial Studies* (forthcoming).

Baker, Malcolm, and Jeffrey Wurgler, 2006. Investor Sentiment and the Cross Section of Stock Returns. *The Journal of Finance* 61, 1645-1680.

Becker C, Ferson W, Myers D, Schill M. 1999. Conditional Market timing with Benchmark investors. *Journal of Financial Economic*s. 52: 119-48.

Berk, J. and R. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269-95.

Boguth, Oliver, et al., 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., and Busse, J.A. (2001). On the timing ability of mutual fund managers. *Journal of Finance* 56, p.1075-1094.

---

nonlinear models do not deliver any significant results.

Brown, S. et al., 1992, Survivorship bias in performance studies. *Review of Financial Studies* 5, 553-580.

Busse, J. 1999. Volatility timing in mutual funds: Evidence from daily returns. *Review of Financial Studies*. 12:1009-41

Cao, Charles, Yong Chen, Bing Liang and Andrew W. Lo, 2011, Can hedge funds time market liquidity? *Working paper*, Virginia Tech.

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

Chen, Y., Ferson, W., and H. Peters, 2010. Measuring the Timing Ability of Fixed Income Funds, *Journal of Financial Economics* 98, 72-89.

Chen, H., Jegadeesh, N. and R. Wermers, 2000, The value of active management: An examination of the stockholdings and trades of fund managers, *Journal of Financial and Quantitative Analysis* 35, 343-368.

Chen, Z., and P. J. Knez. 1996. "Portfolio performance Measurement: theory and applications." *Review of Financial Studies*, 9: 511-556.

Chow, Gregory C. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions". *Econometrica* **28** (3): 591–605.

Christopherson, Jon A., Wayne Ferson and Debra A. Glassman. 1998. "Conditioning manager alpha on economic information: Another look at the persistence of performance." *Review of Financial Studies*, 11: 111-142.

Cochrane, J.H, 1996, A Cross-Sectional Test of an Investment-Based Asset Pricing Model, *Journal of Political Economy* 104(3), 572-621.

Cremers, M. and A. Petajisto, 2009, How active is your mutual fund manager? A new measure that predicts performance. *Review of Financial Studies* 22, 3329-65.

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

De Souza, Andre, and Anthony W. Lynch, 2012. Does mutual fund performance vary over the business cycle? working paper 18137, National Bureau of Economic Research.

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

Elton, E.J. and Gruber, M.J. and Blake, C.R., 2001. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. *Journal of Finance* 56: 2415-2430.

Elton, Edwin J., et al., 2010, The effect of holdings data frequency on conclusions about mutual fund behavior. *Journal of Banking & Finance* 34: 912-922.

Elton, E.J. and Gruber, M.J. and Blake, C.R., 2002, Does mutual fund size matter? The relationship between size and performance, *Review of Asset Pricing Studies* 1, 31-56.

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

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus Skill in the Cross- Section of Mutual Fund Returns. *Journal of Finance* 65, 1915-1947.

Fama, E.F., and K.R.French. 1996. "Multifactor explanations of asset pricing anomalies." *Journal of Finance*, 51: 55-87.

Farnsworth H. K., Ferson W., Jackson D., Todd S., 2002. Performance evaluation with stochastic discount factors. Journal of Business 75: 473-504

French, Kenneth R., 2008, Presidential address: The cost of active investing. *Journal of Finance* 63, 1537-1573.

Ferson, W. 2012, Investment Performance: A review and synthesis, the *Handbook of Economics and Finance*, edited by George Constantinides, Milton Harris and Rene Stulz (forthcoming).

Ferson, W., 1995, Theory and Empirical Testing of Asset Pricing Models, Chapter 5 in *Finance, Handbooks in Operations Research and Management Science*, by Jarrow, Maksimovic and Ziemba (editors), Elsevier, 145-200.

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

Ferson, W. and Min S. Kim, 2012, The factor Stucture of Mutual Fund Flows, *International Journal of Portfolio Analysis and Management* (forthcoming).

Ferson, W. and Jerchern Lin, 2011, Alpha and performance measurement: The effects of investor heterogeneity, *working paper*, University of Southern California.

Ferson, W. and Meijun Qian, 2004, Conditional Performance Evaluation Revisited. Research Foundation Monograph of the CFA Institute (formerly, AIMR) ISBN 0-943205-69-7, 84 pages.

Ferson, W., and Rudi Schadt. 1996. "Measuring Fund Strategy and Performance in Changing Economic Conditions." *Journal of Finance*, 51: 425-462.

Ferson, W. and V. Warther, 1996, Evaluating Fund Performance in a Dynamic Market. *Financial Analysts Journal* 52, no. 6, pp.20-28.

Fleming, J., C. Kirby and B. Ostdiek, 2001, The economic value of volatility timing, *Journal of Finance* 56, 329-352.

Glode, Vincent, 2011, Why mutual funds "underperform". *Journal of Financial Economics* 99: 546-559.

Glosten, Lawrence, and Ravi Jagannathan. 1994. "A contingent claims approach to performance evaluation." *Journal of Empirical Finance*, 1: 133-166.

Graham, J. and C.R. Harvey, 1996, Market timing ability and volatility implied in investment newsletters' asset allocation recommendations," *Journal of Financial Economics* 42, 397-421.

Grant, D. 1977. "Portfolio performance and the 'cost' of timing decisions." *Journal of Finance*, 32: 837-846.

Grinblatt M, Titman S. 1989. Mutual fund performance: an analysis of quarterly portfolio holdings. *Journal of Business* 62: 393-416.

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

Han, Yufeng, 2006, Asset allocation with a high-dimensional latent factor stochastic volatility model, *Review of Financial Studies* 19, 237-271.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029-1054.

Holmes, K., and R. Faff, 2004, Stability, asymmetry and seasonality of fund performance: An analysis of Australian multi-sector managed funds, *Journal of Business Finance and Accounting* 31, 539-578.

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

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

Kacperczyk, M., C. Sialm and L. Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *Journal of Finance*, *60*(4), 1983-2011.

Kacperczyk, M.. C. Sialm and L. Zheng, 2008, Unobserved actions of mutual funds. *Review* *of Financial Studies* 21, 2379-416.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2012, Time-varying fund manager skill, working paper, New York University.

Khang, Minjeong, 2012, Do hedge fund managers possess Timing and Selectivity Skill? Evidence from stock holdings, working paper, Arizona State University.

Kim, M.S., 2011, Changes in mutual fund flows and managerial incentives, *working paper*, University of New South Wales.

Kim, S. and F. In, 2012, False discoveries in volatility timing of mutual funds, *Journal of Banking and Finance* 36, 2083-2094.

Kosowski, Robert, 2011, Do mutual funds perform when it matters most to investors? US Mutual fund performance and risk in recessions and expansions. *Quarterly Journal of Finance* 1: 607-664.

Lakonishok, J., A. Shleifer, and R. Vishny, 1992, "The Structure and Performance of the Money Management Industry," Brookings Papers: Microeconomics, 339–391.

Lehmann, B.N. and D. Modest, 1987, Mutual Fund Performance Evaluation: a Comparison of Benchmarks and Benchmark Comparisons, *Journal of Finance.*

Merton, R. C., 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867-887.

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

Moskowitz, Tobias J., 2002, Discussion. *Journal of Finance* 55: 1695-1703.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica,* 703-708.

Sensoy, B.A., 2009. Performance evaluation and self-designated benchmark indexes in the mutual fund industry. Journal *of Financial Economics* 92: 25-39.

Sharpe WF. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*. 19: 425-42.

Shanken, Jay, 1990, Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, *45*(1), 99-120

Solomon, D., E. Soltes and D. Sosyura, 2012, Winners in the spotlight: Media coverage of mutual fund holdings as drivers of flows, working paper, University of Southern California.

Treynor, Jack, and Kay 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: 1655-1703.

Wang, Junbo, 2013, The statistical properties of holdings-based measures of fund performance, working paper, University of Southern California.

Table 1

Summary statistics of mutual fund characteristics, 1984-2010, are shown for three fund categories (US equity, Asset Allocation, Balanced). The statistics are calculated as the averages over time of the quarterly characteristics for each fund.

		Total	Min	Mean	Median	Max	Std Dev
Number of distinct mutual funds	**All**	2879					
	AssetAllocation	171					
	Balanced	255					
	USEquity	2608					
Number of fund-quarter observations	**All**	98946					
	AssetAllocation	3406					
	Balanced	7357					
	USEquity	88183					
Average TNA (total net assets), $millions	**All**		291	1133	1179	2212	522
	AssetAllocation		295	884	845	1646	300
	Balanced		296	1332	1309	2367	457
	USEquity		291	1158	1185	2412	557
Average Turnover ratio (%) per year	**All**		70.86	87.82	87.69	121.44	10.71
	AssetAllocation		71.96	104.69	103.39	190.43	19.83
	Balanced		71.9	88.02	88.2	118.63	9.3
	USEquity		69.94	87.51	86.71	124.58	11.37
Average Expense ratio (%) per year	**All**		0.92	1.18	1.21	1.41	0.13
	AssetAllocation		0.82	1.25	1.25	1.41	0.11
	Balanced		0.88	1.1	1.08	1.33	0.12
	USEquity		0.92	1.19	1.23	1.41	0.13
Average Proportion invested in stocks, (%)	**All**		78.03	85.91	87.02	94.43	4.9
	AssetAllocation		49.52	62.60	64.78	73.64	6.54
	Balanced		50.28	58.33	59.10	72.87	4.06
	USEquity		79.34	88.76	92.20	96.48	6.11
Average Proportion invested in cash, (%)	**All**		3.1	7.91	6.43	13.95	3.4
	AssetAllocation		6.90	10.36	10.43	18.62	1.86
	Balanced		0.59	6.63	6.30	14.78	2.72
	USEquity		2.95	7.78	6.14	13.95	3.52
Average Proportion invested in bonds, (%)	**All**		2.36	6.18	6.39	9.43	1.81
	AssetAllocation		7.73	27.04	24.08	42.57	7.02
	Balanced		26.48	35.04	34.94	39.73	2.46
	USEquity		0.37	3.46	1.67	7.87	2.67

Table 2
Performance of funds sorted on volatility-related Activity.

This table reports estimates of alpha and its decomposition into market timing, volatility timing, and selectivity. The Carhart 4 factors define the benchmark. The sample covers January of 1998 through December of 2010. The GMM with a Newey-West covariance matrix with three lags is used for estimation. Funds are sorted according to prior estimates of the coefficients of regression (10) in the text as proxies for the likelihood of volatility related active behavior. The average of the sorting variables is shown in the right hand columns.

Panel A: Sorting on λ_{p1} (volatility reaction in reported returns)

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	λ_{p1}
decile 1	**est**	48	0.0048	-0.002	0.0029	0.0036	0.0065	-24.239
(lowest)	**t_stat**		1.0052	-0.511	1.0006	2.0401	2.156	-9.3809
decile 2	**est**	48	0.0033	-0.002	0.0018	0.0028	0.0046	-11.32
	t_stat		0.6744	-0.354	0.6847	1.358	1.422	-8.8965
decile 3	**est**	48	0.0029	-8E-04	0.0021	0.0018	0.0039	-7.1435
	t_stat		0.7715	-0.226	1.0128	1.1655	1.5302	-8.0707
decile 4	**est**	48	0.0023	-6E-04	0.0017	0.0007	0.0025	-4.1527
	t_stat		0.7299	-0.21	1.1285	0.4263	1.0378	-6.3895
decile 5	**est**	48	0.0017	-2E-04	0.0015	0.0004	0.002	-1.6658
	t_stat		0.4986	-0.067	1.0758	0.2414	0.8928	-3.116
decile 6	**est**	48	0.0028	-6E-04	0.0022	-2E-04	0.002	0.712
	t_stat		1.0956	-0.264	1.5542	-0.14	0.9777	1.253
decile 7	**est**	48	0.0012	0.0002	0.0013	-4E-04	0.0009	3.2749
	t_stat		0.4606	0.0725	0.869	-0.23	0.3852	4.3901
decile 8	**est**	48	0.0009	0.0006	0.0015	-4E-04	0.0011	6.2652
	t_stat		0.3448	0.3531	0.7853	-0.243	0.473	5.9091
decile 9	**est**	48	0.0015	0.0007	0.0022	-0.002	0.0005	10.447
	t_stat		0.7162	0.3596	1.0504	-1.049	0.1833	6.633
decile 10	**est**	48	-0.005	0.0035	-0.002	-0.003	-0.005	22.374
(highest)	**t_stat**		-0.948	0.6909	-0.687	-1.298	-1.47	7.3349
decile 10	**est**	48	-0.01	0.0054	-0.005	-0.006	-0.011	46.614
- decile1	**t_stat**		-1.172	0.678	-1.086	-2.661	-2.393	8.5762

Panel B: Sorting on λ_{p2} (volatility timing in reported returns)

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	λ_{p2}
decile 1 (lowest)	**est**	48	0.0012	0.0017	0.0029	-0.003	0.0001	-231.9
	t_stat		0.2539	0.4983	1.1078	-1.478	0.0367	-3.433
decile 2	**est**	48	0.0016	0.001	0.0027	-0.002	0.0004	-110.9
	t_stat		0.3972	0.3057	1.1087	-1.562	0.1702	-3.551
decile 3	**est**	48	0.003	-4E-04	0.0026	-0.001	0.0012	-66.91
	t_stat		0.8772	-0.128	1.3831	-1.068	0.6203	-3.619
decile 4	**est**	48	0.0019	-3E-04	0.0016	-0.001	0.0005	-35.13
	t_stat		0.5726	-0.102	1.0591	-0.728	0.2868	-3.629
decile 5	**est**	48	0.0035	-8E-04	0.0026	-9E-04	0.0017	-8.621
	t_stat		1.1635	-0.309	1.7668	-0.66	0.9026	-2.783
decile 6	**est**	48	0.0017	0.0006	0.0023	0.0007	0.003	16.231
	t_stat		0.7817	0.2994	1.7376	0.4481	1.374	3.2255
decile 7	**est**	48	0.0019	-8E-04	0.0011	0.0017	0.0027	42.071
	t_stat		0.6471	-0.307	0.7367	0.8583	1.0224	3.7152
decile 8	**est**	48	0.0034	-0.002	0.0018	0.002	0.0038	71.396
	t_stat		1.1377	-0.86	0.9696	1.0905	1.1699	3.877
decile 9	**est**	48	-6E-04	-6E-04	-0.001	0.004	0.0027	110.98
	t_stat		-0.308	-0.476	-0.911	1.6741	0.8818	3.9717
decile 10 (highest)	**est**	48	-0.002	0.0006	-8E-04	0.0039	0.0031	220.92
	t_stat		-0.32	0.1643	-0.362	1.1354	0.6922	3.9514
decile 10 - decile1	**est**	48	-0.003	-0.001	-0.004	0.0067	0.003	452.77
	t_stat		-0.388	-0.191	-1.01	1.7222	0.648	3.7

Panel C: Sorting on λ_{p3} (volatility change timing in reported returns)

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	λ_{p3}
decile 1	**est**	48	-0.011	0.0055	-0.005	0.0042	-0.001	-50.11
(lowest)	**t_stat**		-1.443	0.8741	-1.514	1.6839	-0.357	-7.624
decile 2	**est**	48	-0.005	0.0031	-0.002	0.0015	-8E-04	-26
	t_stat		-1.123	0.8846	-0.905	0.6943	-0.266	-7.054
decile 3	**est**	48	-0.001	0.0012	0.0003	0.0018	0.0021	-17
	t_stat		-0.271	0.4533	0.1391	1.0068	0.8447	-6.863
decile 4	**est**	48	0.0007	-1E-04	0.0006	0.0011	0.0018	-10.58
	t_stat		0.1965	-0.028	0.3226	0.7043	0.7067	-6.29
decile 5	**est**	48	0.0019	-5E-04	0.0013	0	0.0013	-5.12
	t_stat		0.514	-0.163	0.8922	0.0141	0.6095	-4.564
decile 6	**est**	48	0.0026	-9E-04	0.0016	0.0002	0.0018	0.0014
	t_stat		0.8078	-0.319	1.1076	0.1176	0.8622	0.0015
decile 7	**est**	48	0.0041	-8E-04	0.0032	-4E-04	0.0029	5.3446
	t_stat		1.4653	-0.35	2.226	-0.299	1.3708	3.8646
decile 8	**est**	48	0.0048	-0.002	0.0027	-3E-04	0.0024	11.572
	t_stat		1.3963	-0.627	1.7508	-0.241	1.1555	5.3054
decile 9	**est**	48	0.0075	-0.002	0.0053	-0.002	0.0034	20.5
	t_stat		1.7442	-0.679	2.171	-1.163	1.1915	5.7401
decile 10	**est**	48	0.0117	-0.004	0.008	-0.003	0.0054	45.526
(highest)	**t_stat**		1.5514	-0.816	1.8453	-1.449	1.16	6.0315
decile 10	**est**	48	0.0225	-0.009	0.0133	-0.0072	0.0066	95.638
- decile1	**t_stat**		1.6199	-0.911	1.8804	-3.138	1.0679	7.19

Table 3

Performance of individual funds, sorting on Factor Model R-squares.

This table reports estimates of alpha and its decomposition into market timing, volatility timing, and selectivity. The Carhart 4 factors define the benchmark. The sample covers January of 1998 through December of 2010. The GMM with a Newey-West covariance matrix with three lags is used for estimation. DGTWcs is the characteristic selectivity measure of Daniel, Grinblatt, Titman and Wermers (1997).

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	DGTWcs
decile 1	**est**	47	0.0043	-0.001	0.0032	0.0037	0.0069	0.0019
(lowest)	**t_stat**		0.8689	-0.256	1.0982	1.1546	1.8112	1.0821
decile 2	**est**	47	0.0052	-0.002	0.0032	0.0033	0.0066	0.002
	t_stat		0.8094	-0.344	0.9416	0.9573	1.382	0.9247
decile 3	**est**	47	0.0033	-2E-04	0.003	0.0029	0.006	0.0015
	t_stat		0.5937	-0.054	1.0124	0.9243	1.4143	0.822
decile 4	**est**	47	0.0035	-0.001	0.0025	0.0018	0.0043	0.0013
	t_stat		0.9386	-0.324	1.1384	0.6737	1.3025	0.7726
decile 5	**est**	47	0.0027	-4E-04	0.0023	0.0016	0.0039	0.0017
	t_stat		1.1378	-0.276	1.2527	0.8376	1.3421	1.0218
decile 6	**est**	47	-0.0009	0.0008	-1E-04	0.0011	0.0009	0.001
	t_stat		-0.5275	0.5333	-0.114	0.8166	0.5342	0.6574
decile 7	**est**	47	-0.0008	0.0005	-3E-04	0.0009	0.0006	0.0006
	t_stat		-0.3733	0.379	-0.18	0.8071	0.4908	0.4897
decile 8	**est**	47	-0.0003	0.0011	0.0009	-2E-04	0.0007	0.0003
	t_stat		-0.1278	0.7623	0.5194	-0.131	0.5569	0.2974
decile 9	**est**	47	0.0005	0.0004	0.001	-0.001	-1E-04	-0.0003
	t_stat		0.1911	0.2681	0.4205	-0.617	-0.122	-0.2594
decile 10	**est**	47	-0.0004	0.0004	0	-0.001	-0.001	-0.0006
(highest)	**t_stat**		-0.1206	0.2019	0.0082	-0.661	-1.049	-0.6476
decile 10	**est**	47	-0.0048	0.0016	-0.003	-0.005	-0.008	-0.0024
- decile1	**t_stat**		-0.7189	0.3525	-0.684	-1.131	-2.138	-1.3067

Table 4
Conditioning Performance on Investor Sentiment

Ex post fund performance is regressed over time on the Baker-Wurgler sentiment index for funds grouped according to their factor model R-squares or on their their Busse volatility reaction coefficients, γ_{1p}, estimated over the past 36 months. The regression slopes are reported (est), along with their Newey-West t-ratios (tstat) and an empirical, two-tailed p-value from a bootstrap simulation under the null that the slope is zero (pval). The betas of stock holdings needed to construct the ex post performance measures are estimated from a rolling window of past one year of daily data with respect to the start of every month.

		Nobs	Level Timing	Volatility Timing	Total Timing	Selectivity	Total Alpha
Panel A: Sorting on Factor Model R-squares							
quintle 1 (lowest)	**est**	141	-0.017	0.002	-0.015	-0.003	-0.018
	tstat		-2.26	0.41	-1.28	-0.71	-1.98
	pval		0.041	0.705	0.264	0.613	0.077
quintle 2	**est**	141	-0.020	0.005	-0.015	-0.002	-0.017
	tstat		-2.60	0.69	-1.23	-0.39	-1.81
	pval		0.03	0.54	0.253	0.808	0.121
quintle 3	**est**	141	-0.024	0.007	-0.017	-0.004	-0.020
	tstat		-3.04	0.83	-1.29	-0.65	-2.25
	pval		0.014	0.442	0.243	0.691	0.049
quintle 4	**est**	141	-0.025	0.007	-0.018	-0.003	-0.022
	tstat		-3.21	0.88	-1.55	-0.67	-2.48
	pval		0.006	0.412	0.149	0.669	0.036
quintle 5 (highest)	**est**	141	-0.025	0.006	-0.019	-0.001	-0.021
	tstat		-3.26	0.89	-1.88	-0.37	-2.47
	pval		0.006	0.432	0.091	0.785	0.02
Panel B: Sorting by Volatility Reaction							
quintle 1 (lowest)	**est**	144	-0.020	0.005	-0.016	-0.004	-0.020
	tstat		-2.45	0.63	-1.19	-0.70	-2.07
	pval		0.033	0.546	0.264	0.598	0.063
quintle 2	**est**	144	-0.020	0.004	-0.016	-0.002	-0.018
	tstat		-2.64	0.66	-1.43	-0.52	-2.13
	pval		0.018	0.529	0.209	0.742	0.064

quintle 3	**est**	144	-0.021	0.005	-0.016	-0.001	-0.017
	tstat		-2.81	0.71	-1.44	-0.36	-2.08
	pval		0.015	0.527	0.217	0.832	0.073
quintle 4	**est**	144	-0.023	0.005	-0.018	-0.002	-0.019
	tstat		-3.11	0.76	-1.56	-0.41	-2.23
	pval		0.007	0.497	0.146	0.792	0.049
quintle 5	**est**	144	-0.027	0.008	-0.020	-0.004	-0.023
(highest)	**tstat**		-3.60	0.89	-1.63	-0.78	-2.57
	pval		0.002	0.44	0.144	0.543	0.024

Table 5

Performance of individual funds, sorting on various predetermined characteristics.

This table reports estimates of alpha and its decomposition into market timing, volatility timing, and selectivity. The Carhart 4 factors define the benchmark. The sample covers January of 1984 through December of 2010 or subsamples as indicated. The GMM with a Newey-West covariance matrix with three lags is used for estimation. DGTWcs is the characteristic selectivity measure of Daniel, Grinblatt, Titman and Wermers (1997).

Panel A:		Sorting on Expense ratios						
		Nobs	**Level Timing**	**Volatility Timing**	**Combined Timing**	**Selectivity**	**Total Alpha**	**DGTWcs**
decile 1	**est**	104	0.0003	-9E-04	-5E-04	0.0009	0.0004	0.0006
(lowest)	**t_stat**		0.2337	-0.825	-0.66	1.0461	0.3794	1.0955
decile 2	**est**	104	0.0002	-3E-04	-1E-04	0.0004	0.0003	0.0002
	t_stat		0.1374	-0.308	-0.135	0.3421	0.2246	0.3529
decile 3	**est**	104	0	-4E-04	-4E-04	-2E-04	-7E-04	0.0005
	t_stat		-0.0171	-0.381	-0.459	-0.226	-0.494	0.6873
decile 4	**est**	104	0.0006	-5E-04	0.0001	0.0004	0.0005	0.0004
	t_stat		0.347	-0.504	0.0659	0.3169	0.2977	0.5466
decile 5	**est**	104	0.001	-6E-04	0.0004	-4E-04	0.0001	0.0006
	t_stat		0.5749	-0.477	0.4216	-0.247	0.0569	0.8911
decile 6	**est**	104	-0.0002	-1E-04	-3E-04	-4E-04	-7E-04	0.0007
	t_stat		-0.1068	-0.074	-0.225	-0.302	-0.429	0.9089
decile 7	**est**	104	0.0013	-0.001	0.0002	-7E-04	-5E-04	0.0004
	t_stat		0.6307	-0.734	0.1277	-0.392	-0.347	0.4917
decile 8	**est**	104	0.0002	-4E-04	-2E-04	-7E-04	-9E-04	0.0003
	t_stat		0.098	-0.321	-0.126	-0.344	-0.506	0.3344
decile 9	**est**	104	0.0004	-6E-04	-2E-04	-7E-04	-9E-04	0.0008
	t_stat		0.2291	-0.581	-0.142	-0.424	-0.557	0.7165
decile 10	**est**	104	0.0006	0	0.0006	-0.001	-8E-04	0.0011
(highest)	**t_stat**		0.3399	0.0147	0.4493	-0.772	-0.556	1.2451
decile 10	**est**	104	0.0002	0.0009	0.0011	-0.002	-0.001	0.0005
- decile1	**t_stat**		0.1368	0.8277	0.9484	-1.708	-1.077	0.6269

Panel B:		Sorting on Active Shares (1990-2007)						
		Nobs	**Level Timing**	**Volatility Timing**	**Combined Timing**	**Selectivity**	**Total Alpha**	**DGTWcs**
decile 1 (lowest)	**est**	71	0.0034	-0.002	0.0012	0.0002	0.0015	3.00E-05
	t_stat		1.4721	-1.267	0.8757	0.1789	1.4025	0.0677
decile 2	**est**	71	0.0031	-0.002	0.001	-6E-04	0.0004	0.0001
	t_stat		1.4189	-1.405	0.7834	-0.438	0.3711	0.0787
decile 3	**est**	71	0.0007	-0.001	-3E-04	0.0003	0	3.5E-05
	t_stat		0.4323	-0.91	-0.346	0.2043	-0.034	0.0404
decile 4	**est**	71	0.0036	-0.003	0.0002	-4E-04	-3E-04	0.0016
	t_stat		1.4114	-1.788	0.1242	-0.307	-0.199	1.1451
decile 5	**est**	71	-0.0001	-8E-04	-8E-04	0.0004	-4E-04	0.0014
	t_stat		-0.0216	-0.395	-0.542	0.2841	-0.271	1.1096
decile 6	**est**	71	-0.0016	0.0006	-0.001	0	-0.001	0.0008
	t_stat		-0.4174	0.1816	-0.509	-0.023	-0.563	0.6608
decile 7	**est**	71	0.0002	-0.001	-8E-04	-0.001	-0.002	0.0017
	t_stat		0.0727	-0.44	-0.381	-0.686	-0.993	1.056
decile 8	**est**	71	-0.0012	-1E-04	-0.001	-8E-04	-0.002	0.0011
	t_stat		-0.2961	-0.026	-0.449	-0.362	-0.688	0.7607
decile 9	**est**	71	0.0015	-0.001	0.0005	-0.002	-0.002	0.0025
	t_stat		0.2983	-0.298	0.1395	-0.981	-0.587	1.3592
decile 10 (highest)	**est**	71	0.0007	0.0005	0.0012	-0.001	-1E-04	0.0014
	t_stat		0.1404	0.128	0.373	-0.591	-0.042	0.9357
decile 10 - decile1	**est**	71	-0.0027	0.0027	-1E-04	-0.002	-0.002	0.0014
	t_stat		-0.5218	0.6218	-0.018	-0.531	-0.493	0.8823

Panel C:		Sorting on Return Gap (1984-2006)						
		Nobs	**Level Timing**	**Volatility Timing**	**Combined Timing**	**Selectivity**	**Total Alpha**	**DGTWcs**
decile 1 (lowest)	**est**	89	0.0001	-0.002	-0.002	0.0023	0	0.0004
	t_stat		0.02	-1.124	-1.45	1.547	0.0121	0.2777
decile 2	**est**	89	-0.0018	0.0002	-0.002	0.0004	-0.001	0.0012
	t_stat		-0.6298	0.0942	-0.802	0.3007	-0.51	1.1879
decile 3	**est**	89	0.0007	-0.003	-0.002	0.0027	0.0007	0.0015
	t_stat		0.2612	-1.391	-1.121	1.7579	0.4235	1.6356
decile 4	**est**	89	0.0007	-0.003	-0.002	0.0015	-6E-04	0.0007
	t_stat		0.262	-1.373	-1.5	1.1555	-0.395	0.8692
decile 5	**est**	89	0.0014	-0.002	-8E-04	0.0006	-2E-04	0.0003
	t_stat		0.5654	-1.102	-0.649	0.5084	-0.113	0.3726
decile 6	**est**	89	-0.0003	-0.002	-0.002	0.0018	-4E-04	0.001
	t_stat		-0.0933	-0.817	-1.43	1.3597	-0.248	1.3069
decile 7	**est**	89	0.0002	-0.002	-0.002	0.002	0.0004	0.0008
	t_stat		0.0708	-0.772	-0.863	1.0718	0.2276	0.9678
decile 8	**est**	89	-0.0004	-0.001	-0.002	0.0014	-4E-04	0.001
	t_stat		-0.1536	-0.646	-1.05	0.984	-0.237	0.9989
decile 9	**est**	89	-0.0003	-0.001	-0.002	0.0013	-2E-04	0.001
	t_stat		-0.1255	-0.712	-0.981	0.8053	-0.088	0.8582
decile 10 (highest)	**est**	89	0.0011	0.0016	0.0027	0.0009	0.0035	0.0014
	t_stat		0.3757	0.8221	0.9965	0.4122	1.2309	0.7667
decile 10 - decile1	**est**	89	0.001	0.0039	0.005	-0.001	0.0035	0.001
	t_stat		0.3861	1.6657	1.6597	-0.91	1.2619	0.9468

Panel D:		Sorting on Turnover						
		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	DGTWcs
quintle 1 (lowest)	**est**	96	-3E-05	-2E-04	-3E-04	0.0006	0.0004	0.0004
	t_stat		-0.0197	-0.225	-0.287	0.4906	0.2525	0.469
quintle 2	**est**	96	-0.001	-4E-04	-0.001	-5E-04	-0.002	-0.0005
	t_stat		-1.0335	-0.939	-1.847	-0.545	-1.938	-0.7873
quintle 3	**est**	96	-0.0022	0.0012	-0.001	0.0007	-3E-04	0.0004
	t_stat		-1.3975	1.331	-0.767	0.5121	-0.168	0.3978
quintle 4	**est**	96	4E-06	-3E-04	-3E-04	-3E-04	-6E-04	0.0002
	t_stat		0.0023	-0.371	-0.263	-0.165	-0.4	0.1921
quintle 5 (highest)	**est**	96	0.002	0.0005	0.0025	-0.001	0.0015	0.0017
	t_stat		0.9782	0.4218	1.6905	-0.473	0.7365	1.2003
quintle 5 - quintle 1	**est**	96	0.002	0.0008	0.0028	-0.002	0.0011	0.0013
	t_stat		0.7643	0.4082	1.6666	-0.815	0.5756	0.7769

Panel E:		Sorting on Turnover (equity-only positions)						
		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	DGTWcs
quintle 1 (lowest)	**est**	96	0.0002	-5E-04	-2E-04	0.0006	0.0003	0.0004
	t_stat		0.1133	-0.34	-0.164	0.3919	0.2099	0.3782
quintle 2	**est**	96	-0.0005	-3E-04	-8E-04	-5E-04	-0.001	-0.0005
	t_stat		-0.4389	-0.63	-0.789	-0.48	-1.381	-0.5835
quintle 3	**est**	96	-0.0019	0.001	-9E-04	0.0008	-1E-04	0.0003
	t_stat		-1.3048	1.5096	-0.672	0.5132	-0.065	0.305
quintle 4	**est**	96	0.0006	-6E-04	1E-05	-4E-04	-3E-04	0.0003
	t_stat		0.3216	-0.616	0.01	-0.187	-0.265	0.2851
quintle 5 (highest)	**est**	96	0.0023	0.0001	0.0025	-9E-04	0.0016	0.0016
	t_stat		1.1253	0.1088	1.5992	-0.37	0.9987	1.0834
quintle 5 - quintle 1	**est**	96	0.0021	0.0006	0.0027	-0.001	0.0013	0.0012
	t_stat		0.7748	0.3024	1.6858	-0.684	0.6203	0.6779