Performance Measurement with Market and Volatility Timing and Selectivity

by
Wayne Ferson
and
Haitao Mo

First draft: September 5, 2011
This revision: February 26, 2012

PRELIMINARY

ABSTRACT:

To measure the investment performance of a portfolio manager who may engage in market timing, it is necessary to consider both market level and volatility timing behavior as well as security selection ability. We develop and implement measures that accommodate all three components. A well specified measure of performance is the sum of the three components of ability. Estimating the measures on active US mutual funds, we find that allowing for market level and volatility timing, there is no evidence of investment ability at the level of broad groups of mutual funds, or when funds are sorted by expense ratio, return gap, active share or turnover. However, sorting by factor model R-squares confirms results in Amihud and Goyenko (2011), where the low R-square 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, Haitao.Mo.2012@marshall.usc.edu. We are grateful to Fabio Moneta for help with some data, to participants at the 2012 Sonoran Finance Conference for helpful discussions, and to our discussant Oliver Boguth.

Performance Measurement with Market and Volatility Timing and Selectivity

PRELIMINARY

February 26, 2012

ABSTRACT:

To measure the investment performance of a portfolio manager who may engage in market timing, it is necessary to consider both market level and volatility timing behavior as well as security selection ability. We develop and implement measures that accommodate all three components. A well specified measure of performance is the sum of the three components of ability. Estimating the measures on active US mutual funds we find that, allowing for market level and volatility timing, there is no evidence of investment ability at the level of broad groups of mutual funds, or when funds are sorted by expense ratio, return gap, active share or turnover. However, sorting by factor model R-squares confirms results in Amihud and Goyenko (2011), where the low R-square 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 only one type of ability exists. This leads to misspecification if both types of behavior may be present. Measures of performance that account for market timing have been developed, but these make very strong assumptions such as "perfect" timing ability of a stylized form, or the validity of a simple options pricing model. (See for example the discussion in Aragon and Ferson, 2008.) Furthermore, measures that do attempt to accommodate timing behavior typically model the ability to time the level of market factors, but not to time market volatility. Busse (1999) documents volatility timing behavior in US mutual funds. If both level and volatility timing behavior are present, models that leave out one of them are misspecified.[1]

This paper develops and implements simple measures of performance that account for both market and volatility timing as well as security selection ability. A well-specified measure of total performance is a weighted sum of the three components. Using our new measures, we revisit the issue of US mutual fund performance.

The primary reason that previous measures of performance have trouble with market timing behavior is the nonlinearity it creates in portfolio returns. Classical

[1] Cao, Chen, Liang and Lo (2011) consider both market and volatility timing for hedge

measures of timing, such as Treynor Mazuy (1966) and Merton and Henriksson (1981) identify timing ability through this nonlinearity. One problem with assessing timing in this way is that there are many other sources of nonlinearity that may be unrelated to market timing (see for example, Chen Ferson and Peters (2010) for a discussion). Second, if timing is left out of a returns-based performance regression, the performance 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.

Our measures use portfolio weights, and are thus related to the rapidly developing literature on holdings-based performance measures that Grinblatt and Titman (1989, 1993) kicked off. We develop the relation of our measure to previous weight-based measures, and point out sources of potential misspecification in those measures that our new measures can avoid. We show that holdings-based performance measures are misspecified if they leave out the volatility timing component. All previous studies, excepting Busse (1999) and Cao et al. (2011) to our knowledge leave out this component. Boguth et al. (2011) suggest that volatility timing may impart substantial biases to estimates of alpha. Aragon and Martin (2011) suggest that hedge funds may actively time volatility. Our measures extend the weight-based performance measures of Grinblatt and Titman (1989, 1993) to accommodate volatility timing and selectivity in a parsimonious way. Only three parameters are needed for each mutual fund, which is much more tractable than returns-based market timing models. This allows us to easily examine models with multiple benchmarks.[2]

---

funds, but their returns-based approach is very different from ours.

[2] Of course, using weights we do not exploit the information in high frequency mutual fund returns, which Bollen and Busse (2001) find helps to detect market timing ability. Ferson

We implement our measures on a sample of US active, open-ended mutual funds. We find that when only level and volatility timing ability are measured in an asset allocation setting, the measures suggest weak negative timing ability that is not statistically significant, a finding consistent with a large literature on timing ability that does not include all the components of performance. When level timing, volatility timing and selectivity ability are examined jointly, we find no evidence of ability at the level of broad groups of funds. Sorting funds on expense ratios, turnover, return gap or active share reveals no significant performance, either. Sorting on factor model regression R-squares, following Amihud and Goyenko (2011), we confirm their finding that the low R-square funds deliver higher alphas and greater selectivity performance. With this exception, the findings seem to confirm the results of recent studies such as Fama and French (2010), who find little evidence of performance based on traditional measures of alpha. Our results also suggest that Busse's (1999) finding of volatility timing behavior is not robust and that previous studies finding ability using weight based measures (e.g. Grinblatt and Titman (1993), Daniel, Grinblatt Titman and Wermers, 1997) may have been biased by omitted variables.

The rest of the paper is organized as follows. Section 2 describes the models and their estimation. Section 3 describes our data. Section 4 presents the main empirical results and Section 5 concludes and offers suggestions for future research.

---

and Khang (2002) examine the power of weight-based approaches and find that the information in the portfolio weights offsets the loss of information in reported fund returns, and that measures using weights can be quite powerful. While we use daily stock return data in some of our analyses, combining these data with high frequency mutual fund data in a mixed-interval data analysis is a good topic for future research. By abstracting from costs, weight-based measures miss the possibility that the ability to trade at low cost may be a form of skill.

2. The Models

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

$$\alpha_p = E(m\, r_p), \qquad (1)$$

where m is the stochastic discount factor 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 (2011) argue that if m is the client's marginal rate of substitution, Equation (1) is the best way to specify a valid performance measure. We assume the SDF is given by popular linear factor models, following Cochrane (1996):

$$m = a - b' r_B, \qquad (2)$$

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 special case is the Capital Asset Pricing Model (CAPM, Sharpe, 1964) where K=1 and a broad market index is the benchmark. We start with equations (1) and (2) in their simple "unconditional" form and discuss conditional versions of the model below.

2.1 A Simple Asset Allocation Model

We start with the simplest model that emphasizes the two kinds of market timing ability in an asset allocation setting. Assume the fund forms a portfolio from the benchmarks with weights, w, set at the beginning of the period before the returns are realized, so the portfolio excess return is $r_p = w' r_B$. In this formulation, the "cash" position invested in the short term Treasury security is $1 - \underline{1}' w$, where $\underline{1}$ is a K-vector of ones. A simple performance measure follows by substitution of Equation (2) into Equation (1):

$$\alpha_p = a\, E(w' r_B) - b\, E(r_B r_B'\, w). \tag{3}$$

The first term of (3) captures market level timing through the covariance between the portfolio weights and the subsequent factor returns. The second term captures "volatility timing," through the relation between the portfolio weights and the second moment matrix of benchmark returns.

The benchmarks have zero alphas in (1) by construction when (2) describes the SDF: $\alpha_B = a\, E(r_B)' - b\, E(r_B r_B') = 0$. We can use this to write the measure in (3) in terms of covariances:

$$\alpha_p = a\, \mathrm{Cov}(w' r_B) - b'\, E\{ [r_B r_B' - E(r_B r_B')]\, w\}, \tag{4}$$

where the notation $\mathrm{Cov}(w' r_B)$ denotes the sum of the covariances between the weights and the corresponding future benchmark excess returns. In our estimation scheme, described below, we use Equation (4).

Equation (4) provides insight about and offers practical advantages over traditional measures. The first term in (4) is essentially the weight-based measure of Grinblatt and Titman (1989, 1993) applied at the "asset allocation" level. Like their measure and its subsequent elaborations (see e.g., Daniel et al. (1997), Wermers, 2000), the measures focus on portfolio weights instead of the reported, after-fees-and-costs returns of the mutual funds. This measures manager ability on a clean, before-cost basis. The performance can then be compared with costs. Traditional returns-based measures of performance mix the after-cost returns of funds with the before-cost returns of the benchmarks, creating an "apples to oranges" comparison.

The analysis here shows that the original Grinblatt and Titman measure is misspecified in the presence of market timing behavior, because it excludes the volatility timing term. This is because their measure is developed under joint normality with homoskedasticity, so an informed manager never gets a signal that second moments will change. We evaluate the empirical impact of the missing timing terms.

Previous market timing measures, 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 benchmark returns, and with multiple benchmarks, the products of the benchmarks. With K benchmarks, there are $2K + (K^2 - K)/2$ coefficients to be estimated. For example with three factors there are 9 slope coefficients plus an intercept in the regression for each fund. This can be a degrees-of-freedom challenge when many mutual funds have short sample histories. In contrast, the second term of (4) uses the K x K second moment matrices of the benchmarks, aggregated through a single market-wide parameter, b, of length K, appropriate for their valuation by the model. This only requires two parameters for each fund; one captures market level timing and one captures volatility timing.[3] 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.

2.2 Security Selection

We now bring security selection ability into the model. Consider a factor model regression for the excess returns of the N underlying securities:

$$r = a + \beta\, r_B + u, \tag{5}$$

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 intercept plus residuals: $v = a + u$. Security selection ability means that the past portfolio weights of a fund are correlated with future values of v. A fund forms a portfolio of the N assets using weights, x, as:

$$r_p = x'r = (x'\beta)r_B + x'v. \tag{6}$$

Note that $w'=x'\beta$ corresponds to the K-vector of asset allocation weights in the simpler market timing model. Our approach is to estimate these using a "bottom up" method and

---

[3] 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 regression, further expanding the number of regressors and using up degrees of freedom.

daily data for the underlying asset returns and benchmarks, similar to Jiang, Yao and Yu (2007) and Elton, Gruber and Blake (2009). Substituting Equation (6) into the definition of alpha we obtain:

$$\alpha_p = a\, E(w' r_B) - b'\, E(r_B r_B'\, w) + E\{(a - b' r_B)\, x'v\}, \quad (7)$$
$$= a\, Cov(w' r_B) - b'\, E\{\, [r_B r_B' - E(r_B r_B')]\, w\} + E\{(a - b' r_B)\, x'v\}.$$

The first two terms of each line in the equations (7) appear in equations (3) and (4) respectively, while the third term captures selectivity ability. The third term says that if the portfolio-weighted average of the idiosyncratic returns of the securities in the fund have a positive covariance with the marginal rate of substitution, they represent performance with positive value.

It helps to the break the last term of Equation (7) into two pieces for interpretation:

$$E\{(a - b' r_B)\, x'v\} = a\, E(x'v) - b'\, E(r_B v' x). \quad (8)$$

The first piece is a classical measure of selectivity; that is, the covariance between the portfolio weights and the subsequent abnormal returns, summed across the securities in the fund. The second term relates to information in the portfolio weights about future second moments and is priced using the coefficient, b, in the same way as the volatility timing ability. This term captures the possibility that managers' superior information reveals that the residuals of the factor model regression (5) may not be uncorrelated with the factors, conditional on their finer information. For example, a manager might have information that a stock is likely to have a large idiosyncratic return if the stock market return is high. The selectivity ability in Equation (8) is zero when the portfolio weights have no information about the future idiosyncratic return, v, and no information about the product, vr_B.

2.3 Relation to Previous Holdings-based Performance Measures

As mentioned above the first term of Equations (7) is essentially the weight-based performance measure first developed by Grinblatt and Titman (1989, 1993) applied at the asset allocation level. To relate our measure more explicitly to the original Grinblatt and Titman measure, substitute in the regression (5) and use $x'\beta = w'$ to see that:

$$Cov(x'r) = Cov(w'r_B) + Cov(x'v). \quad (9)$$

This expression shows that the Grinblatt Titman measure on the left-hand side of (9) leaves out the two terms related to information about the second moments that appear in our measure in (7). There is no volatility timing term, and the second component of selectivity in Equation (8) is missing. If information about time-varying second moments, as captured in our measure, is important, then the original measure is misspecified for these two reasons.

A number of popular performance measures in the literature essentially add and subtract pieces to the Grinblatt and Titman measure, which are interpreted as components of the total performance. A well-cited example is Daniel, Grinblatt, Titman and Wermers (DGTW, 1997). In this example each security *i* held in a fund gets its own benchmark return, R_t^{bi} at each period, t. In addition, the fund is assigned a set of benchmark weights equal to its actual holdings reported k periods before: $x_{i,t-k}$. The DGTW measure is:

$$DGTW_{t+1} = \Sigma_i\, x_{it}\, (R_{i,t+1} - R_{t+1}^{bi}) + \Sigma_i\, (x_{it}\, R_{t+1}^{bi} - x_{i,t-k} R_{t+1}^{bi(t-k)}) + \Sigma_i\, x_{i,t-k}\, R_{t+1}^{bi(t-k)}\,, \qquad (10)$$

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 "selectivity," the second term as "characteristic timing" and the third as the return attributed to the style exposure. The characteristic timing term is similar to our term that measures level timing. The DGTW measure does not capture volatility timing, so it will be misspecified if volatility timing is important.

If we take the security specific benchmark, R_{t+1}^{bi} in (10), as the analogue to the systematic component of return in our Equation (6), then the first term measuring selectivity in the DGTW measure is analogous to the first component of our selectivity term in Equation (8). We will compare the two measures below. Our equation (8) shows that the DGTW measure is misspecified as a measure of selectivity if the higher order effects, $E(r_B v'x)$ are important.

The original Grinblatt and Titman measure and the DGTW measure use unconditional covariances, and are also misspecified if conditional covariances given public information are important, as shown by Ferson and Khang (2002). 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.

2.4 Estimation

We estimate the market-wide parameters a and b through the short-term Treasury return and the excess return of the benchmarks, as shown in Equations (11a-11b) below. For each fund we estimate a market timing component, denoted below 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:

$$\varepsilon_1 = (a - b' r_B) r_B \tag{11a}$$

$$\varepsilon_2 = (a - b' r_B') R_f - 1 \tag{11b}$$

$$\varepsilon_3 = r_B - \mu_B \tag{11c}$$

$$\varepsilon_4 = \alpha_m - a(r_B - \mu_B)'w \tag{11d}$$

$$\varepsilon_5 = \alpha_\sigma + b'(r_B r_B')w - a\, \mu_B' w \tag{11e}$$

$$\varepsilon_6 = \alpha_S - [(a - b' r_B)\, v'x]. \tag{11f}$$

In Equation (11e) 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 (11a) and (11b), no additional restrictions are imposed in using this condition. In Equation (11f), we use $v = r - W r_B$, where W 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.[4] 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 with the delta method to get standard errors for the total performance measures.

[4] 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

The GMM system (11) is exactly identified and has a block diagonal structure with respect to the fund-specific performance parameters, which is particularly convenient for our application. 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 data for 1984-2010 from the Center for Research in Security Prices Mutual Fund database. We exclude fixed income, international, money market, sector and index funds,[5] focusing on active, US equity funds. 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). 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 data set. We combine multiple share classes for a fund, focusing on the TNA-weighted aggregate share class.

We first focus our analysis on portfolios of funds. We study US open-ended Equity, Asset Allocation and Balanced funds. These broad groups are determined using the

---

with daily data, this caveat should be kept in mind when interpreting our empirical results.
[5] We identify and remove index funds both by Lipper objective codes (SP, SPSP) and by searching the funds' names with key word "index."

investment objective codes from CRSP.[6] 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 WFICN links.

We group funds according to several attributes, including the expense ratio, fund size (TNA), age, turnover, return gap, active share and factor model regression R-squares. These are described more fully below.

We use daily returns data to estimate "bottom up" 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.

4. Empirical Results

4.1 Market and Volatility Timing Using Asset Allocation Weights

We start with a simple case, where the CAPM is the benchmark model and we focus

---

[6] 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 those with policy code: Bal; Weisenberger objective code BAL; SI objective code BAL or Lippoer objective codes B, BT, MTAC, MTAG, MT or AM.

on timing ability. Thus, we estimate only the equations (11a) – (11e). For this exercise we take the asset allocation weights from CRSP, which reports the percentage holdings in stocks, bond and cash on an annual or quarterly basis. Our analysis is based on a quarterly holding period. Although some forms of ability may reveal themselves at longer horizons, it seems likely that at least some of the benefits to smart position would accrue during the first quarter.

Table 1 presents our estimates of performance and its decomposition into market timing and volatility timing for each fund group: Asset Allocation, Balanced and the full active US Equity sample. The estimate of the market-wide parameter a is strongly statistically significant, at 1.03, and the estimate of the parameter b is 2.42 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. The values seem to imply economically reasonable magnitudes.

The performance estimates in Table 1 suggest insignificant "negative" market level timing ability for the combined active equity sample, and insignificant positive ability for the Asset Allocation group, which is similar to many studies of unconditional market timing ability. Market volatility timing ability is small and not statistically significant. The combined performance measure is also small, positive for the Asset Allocation group at 6 basis points per quarter, and not statistically significant, with a t-ratio of 1.81. We find similar results when we further break each group of funds into thirds on the basis of their expense ratios, turnover, size or age (these results are not tabulated).

Possibly, the market timing model using only the market benchmark is misspecified, as funds might consider allocations to bonds, stocks and cash. Since stock and bond returns are correlated, leaving out the bonds could bias the market timing results. Our second example is a two-factor 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. The weight in bonds includes convertible bonds, corporate bonds, municipal bonds and government bonds. The weights are normalized to sum to one minus the reported holdings in cash. This allows us to see how excluding a bond market factor, as most studies of market timing have done, might affect the findings.

The results for the two-factor asset allocation model are summarized in Table 2. The point estimate of the parameter *a* is 1.19 and the point estimate of *b* for the market index is 2.47, both very similar to Table 1. The value of the *b* parameter for the bond index is 16.76. All of these coefficients are statistically significant. 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 and the variance of the bond factor, the latter being a relatively small number. This results in a scaled-up value of the parameter *b*.

The performance estimates in Table 2 suggest small but insignificant "negative" market level timing and small positive volatility timing ability, the latter consistent with Busse (1999). The volatility timing coefficients are at the margins of statistical significance, with t-ratios of 2.02 or less. The combined performance measure is close to zero, at 9 basis points per quarter or less. We again find similar results when we further break each group of funds into thirds on the basis of their expense ratios, turnover, size or age (these results are not tabulated). Overall, this section shows that or framework reproduces results essentially similar to what previous studies of unconditional timing ability have found through other methods.

4.2 Selectivity and Multiple Benchmarks

Possibly, the weak evidence for investment ability at the asset allocation level is related to the use of CRSP data for the asset allocation weights. Now we bring in the holdings data and measure the portfolio weights, x, held by the fund in individual stocks. We combine this with the CRSP data on holdings in bonds and normalize the weights to sum to one minus the CRSP reported holdings in cash. The asset allocation weights for the benchmarks are now derived "bottom up" from the individual stocks' betas, estimated using daily data over the full available sample for each stock. We consider two multifactor benchmarks: the FF3 factors and the Carhart (1997) 4-factor model. The results are summarized in Tables 3 and 4.

Table 3 presents results using the FF 3 factor benchmark. The point estimates of the parameter *a* and the *b* coefficient for the market factor are similar to the previous cases, but the *b* parameter is no longer statistically significant. The b coefficients on the HML and SMB factors are also not significant.

The timing performance results in Table 3 are essentially similar to the results using only the asset allocation weights. The selectivity ability term is small and statistically insignificant. The total alphas in Table 3 are economically small, at 14 basis points per quarter. The standard errors say that the alphas are also reliably close to zero. For example, for the US equity sample, the estimate of alpha is 11 basis points and the standard error is 14 basis points per quarter, so a two-standard error confidence band covers (-17, +39) basis point per quarter. These values are reliably smaller than the average of expense ratios plus trading costs. We find similar results when we further break each group of funds into thirds on the basis of their expense ratios, turnover, size and age (these results are not tabulated).

Table 4 presents results using the Carhart 4-factor benchmark. All four of the b coefficient estimates are positive and the coefficients for the HML and SMB factors are now statistically significant. The performance results are similar to those for the FF 3 factor model.

4.3 Conditioning Information

Previous studies find that performance inferences about timing and overall performance can be sensitive to the effects of public information. In particular, Ferson and Schadt (1996) and Becker et al. (1999) find that market timing ability looks better in models that account for public information. Consistent with those studies, we find weak negative timing ability in the unconditional models, which combines to produce insignificant total alphas. Perhaps, conditional models can produce a different result. This section considers conditional versions of the models.

4.4 Conditional Models

The conditional models follow studies such as Cochrane (1996) and assume that the parameters a and b are linear functions of the lagged instruments, Z, and that the conditional means of the benchmark excess returns are linear functions of Z.[7] Thus, a and b are replaced by the linear functions a(Z) and b(Z), and μ_B is replaced by $\delta_B Z$ in Equations

[7] 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

(11c-e), where δ_B is a K x L matrix of parameters and L is the number of lagged instruments in Z. The modified equations (11a-c) are multiplied by each element of Z. The GMM with a Newey-West covariance matrix using three lags is used in estimation of the standard errors, and the system is exactly identified.[8]

Table 5 reports results for conditional market timing in the CAPM. The SDF coefficients aj and bj, j=0,…,4 are reported in the following order: Intercept, dividend yield, Tbill yield, TERM and DEF. 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, while the we do not reject the hypothesis that a(Z) is a constant function over time.

The performance results in Table 5 indicate insignificant overall market timing ability and the point estimates of both components of timing ability are individually insignificant. This is consistent with conditional performance evaluation studies such as Ferson and Schadt (1996) and Becker et al. (1999). Moving to a conditional version of the CAPM removes any evidence of negative timing ability. The combined ability is reliably small. For example, for the Asset Allocation funds a two-standard error confidence band covers (-3, +9) basis points per quarter. We find similar results when we further break each group of funds into thirds on the basis of their expense ratios, turnover, size or age (these results are not tabulated).

We also examine a conditional asset allocation model with two factors: the market index and the bond index, similar to Table 2. There is less evidence of time-varying SDF

---

nonlinear models do not deliver any significant results.

[8] We do not allow the components of ability to vary over time in this draft, focusing on the expected conditional alphas. It might be interesting to allow for time varying conditional

coefficients in this model, consistent with the less-significant b coefficients in Table 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 small.

In Table 6 we estimate a conditional version of the FF3 factor model and examine all three components of performance using the Thompson holdings data. The b-parameters for the market index are significantly-related to the lagged dividend yield, Tbill yield and Default spread, and we would also reject the hypothesis that the b-coefficient for the SMB factor is constant over time. The intercept and HML coefficients are not significantly time varying. The performance results indicate that the combined timing ability, considering both level and volatility timing, is small and positive but insignificant at the fund group level. The selectivity component is also insignificant. When we break each group of funds into thirds on the basis of their expense ratios, turnover, size or age we find similar results spread across virtually all of the groups.

We also estimate a conditional version of the Carhart 4-factor model. The b-parameter for the market index is significantly-related to the lagged dividend yield and Tbill yield, but none of the other b-coefficients nor the *a* coefficient are significantly time varying. The combined timing ability is numerically close to zero and statistically insignificant. The selectivity component and total alpha are also insignificant. When we break each group of funds into thirds on the basis of their expense ratios, turnover, size or age we find similar results spread across virtually all of the groups.

In summary, the conditional models confirm previous evidence, using other

---

alphas along the lines of Christopherson et al. (1998).

approaches. Allowing for both level and volatility timing behavior, the evidence for selectivity ability and positive overall performance is nil at the aggregate group levels. The economic magnitude of the performance is also reliably small.

4.5. Sorting Funds on Characteristics

It may not be surprising to find no significant performance at the level of large groups of funds, but there may be some individual funds with significant performance. Previous studies have identified fund characteristics associated with differences in performance using other performance measures. We consider several characteristics in this section. These include the expense ratio, turnover, active share, return gap and factor model regression R-squares. The results are summarized in tables 7 and 8. We include for comparison purposes, the "characteristic selectivity" measure from Daniel, Grinblatt, Titman and Wermers (1997), which is the first term of our Equation (10), and denoted as DGTWcs.

In panel A of Table 7 we sort funds into equally-weighted quintiles according to their most recently reported expense ratios. Consistent with previous studies, the DGTWcs measures are positive, with t-ratios larger than 1.9 for four of the five quintiles. The measures are higher for the higher expense ratio funds, attaining about 1% per year for the highest expense ratio group, but the difference across the quintiles is not statistically significant. Grinblatt and Titman (1993), Wermers (2000) and others find positive DGTWcs measures and suggest that US equity funds have positive selectivity ability on a before cost basis. Our new measures produce a different result. The combined performance alphas are positive, but less than 10 basis points per quarter for each group and the t-ratios are small.

In fact, all the components of performance, including the selectivity measures, are small and insignificant. The patterns suggest that the DGTWcs measures are likely biased by the omission of the other components. Six of the ten point estimates of timing ability are negative, and these combine with the selectivity components to produce small alphas in our models.

Amihud and Goyenko (2011) find that when funds are sorted according to the regression R-squares of the returns on standard factor models, the funds with lower R-squares have higher subsequent performance. Panel B of Table 7 sorts funds in this way and examines our measures along with the DGTWcs measure. Here we find our strongest evidence of ability. Consistent with Amihud and Goyenko, the DGTWcs measure is larger for the lower R-square funds. Our selectivity measure also displays this pattern, and with economically similar magnitudes. The difference across the deciles is about 2.2% per year, but does not attain statistical significance. Our overall alpha measures vary substantially with the R-squares, ranging from zero to 87 basis points per quarter, and the difference between the high and low-R-square group sports a t-ratio over two. This difference, about 3.5% per year, is likely economically significant.

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 excess returns differ significantly, by about 2.5-3% per year across quintiles, and the more active funds deliver higher future returns. Petajisto graciously provides data for the active shares on his web page. Panel C of Table 7 sorts funds by the active shares and examines our performance measures along with the DGTWcs measure. The DGTWcs measures increase with the active shares and the levels are positive, but the differences across the groups are not statistically significant. Using our performance measures we find small and statistically insignificant performance, with no clear relation to the active shares. Both the selectivity measures and the overall alphas are precisely close to zero.

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 (2011) finds that the flow-performance relation in mutual funds attenuates after the year 2000, which could be related aggregate 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 used data that do not cover the period after 2000. Perhaps, the evidence for investment ability has changed in the more recent data. In Table 8 we repeat some of the analyses from Table 7, using the restricted 1984-1999 sample period. We also include analyses where we sort funds on return gap and turnover. We use daily fund returns to estimate the factor model R-squares. Since CRSP daily fund return data starts from 1999, we do not include the R-square measure in Table 8.

In panel A of Table 8 funds are sorted into deciles on the basis of their most recently-reported expense ratios, and the findings are similar to those in Table 7. In Panel B we sort funds on their active shares and the results are similar to those in Table 7, except our overall alpha estimates produce negative t-ratios larger than two in a couple of cases. Thus, we find no evidence that the findings of no performance using our measures are driven by the post 2000 data.

Kacperczyk, Sialm and Zheng (2007) 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. Sorting by return gap in Panel C of Table 8, the DGTWcs measures are again positive and many appear statistically significant, but there are no clear patterns across the deciles. Using our performance measures we find marginally significant positive ability in a few cases, but no significant relation of performance to the return gap.[9]

Wermers (2000) finds finds a positive relation between turnover and a fund's DGTWcs measure. We sort our sample of funds into quintiles based on turnover each year and estimate our models on each quintile portfolio. We find in Panel D of Table 8 like Wermers, that the DGTWcs measure is higher for the higher turnover groups. The differences are marginally statistically significant. However, using our measures on either the Fama-French three-factor or the Carhart four-factor benchmarks, we find no significant relation of any performance measure to the turnover group. The overall alphas are economically small and precisely close to zero.

4.6 Precision and Power

Our findings of no significant performance naturally raise the question of statistical power. We are working on simulations to address that issue. However, it seems that the precision of the estimates is high, suggesting that we would have detected economically

[9] The original Kacperczyk et al. (2007) return gap data are available on the RFS web site. We redo the analysis from 1984-2006 using these to sort the funds. We find that half of the DGTWcs measures for the deciles sport t-ratios larger than two and the measures range from 10 to 33 basis points per quarter but the difference between the highest and lowest return gap decile is not statistically significant. Our selectivity measures have a similar range and two of the t-ratios are lager than two. The only marginally statistically significant difference between the high and low return-gap deciles is for volatility timing, with a t-ratio of 1.9.

significant performance if it existed. For example, in tables 3-6 the average standard error of the alpha estimates is about 36 basis points per year, so performance of 1% per year would earn a t-ratio of about three.[10] The small t-ratios correspond to point estimates of performance that are economically small.

4.7. Additional Tests

We conduct some additional experiments to replicate some of the earlier work that found positive performance using weight-based performance measures. 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, no doubt resulting in noisy estimates, but if we require that a fund survive for a longer period, the survival selection induced bias can create spurious persistence (e.g. Brown et al. 1992). We find no evidence of persistence in any of our measures.

Studies such as Ferson and Qian (2004), Moskowitz (2000), Kowsowski (2011), Glode (2010) and Kacperczyk, Nieuwerburgh and Veldcamp (2011) suggest that fund performance may vary over the state of the business cycle, possibly with stronger performance in recession periods. However, DeSouza and Lynch (2011) criticize some of these studies for using NBER reference cycles, which are only available *ex post*, to condition the analysis. They find that the evidence for business cycle variation in performance weakens substantially or disappears when ex ante conditioning variables are used. We estimate a probit model for the likelihood of a recession and break the sample up into high,

[10] This is much more precision than is available with returns-based alphas, which are typically on the

medium and low recession probability subsamples. While we find a higher incidence of t-ratios above 2.0 that would be expected by chance in the recession-state-conditioned performance measures (nine out of 45 cases), but there is no clear pattern in their incidence.

5. Conclusions

To measure the total investment performance of a portfolio manager who may engage in market timing, it is necessary to consider both level and volatility timing behavior as well as selectivity ability. This paper develops and implements simple measures of performance that account for all three components and shows that a well-specified measure of total performance is a weighted sum of the three components. Estimating the measures on US mutual funds, we find some evidence consistent with previous studies, such as weak negative market level timing and weak positive volatility timing, when some of the terms are omitted from the model as they were in earlier studies. But, when all three terms are present we find virtually no evidence of ability at the fund group level or for individual funds sorted according to various fund characteristics that previous research has found to be related to performance. The exception is the regression R-squared from regressing the funds' returns on common factors, where we confirm a finding from Amihud and Goyenko (2011) that the low R-square funds have stronger performance. With this exception, our results seem to confirm the results of recent studies such as Fama and French (2011) there is little evidence of ability. The results also suggest that Busse's (1999) finding of volatility timing behavior is not robust, and that previous studies finding ability using weight based measures may have been biased by omitted

---

order of 25 basis points per month.

variables. If funds' investment ability is convincingly neutral on a before cost basis, it would suggest that investors are wasting money roughly equal to the costs of fund management, calculated by French (2008) to be about 2/3 of one percent per year.

References:

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

Aragon G, Ferson W. 2007. Portfolio Performance Evaluation. *Foundations and Trends in Finance* 2:2: 1-111

Aragon, G. and S. Martin, 2011, Journal of Financial Economics,

Amihud, Y. and R. Goyenko, 2011, Mutual fund R^2 as a predictor of performance, working paper, New York University.

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

Boguth, Oliver, M. Carlson, A. Fisher and M. Simutin, 2011, Journal of Financial Economics,

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

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*.

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

Chevalier, J., and L. Ellison, 1997, Risk Taking by Mutual Funds as a Response to Incentives, *Journal of Political Economy* 105, 1167-1200.

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.

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

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, E. and Gruber, M. and Blake, C., Forthcoming. An examination of mutual fund timing ability using monthly holding data. *Review of Finance*.

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

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

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

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

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

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.

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

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

Koski, J. and J. Pontiff, 1999, How Are Derivatives Used? Evidence from the Mutual Fund Industry," *Journal of Finance* 54(2), 791-816.

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

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.

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.

Table 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 (11).

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 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 a and b in the system (11).

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	**Volatiliy Timing**	**Sum**
AssetAllocation	**Est**	33	-7E-04	0.0008	0.0001
	t_stat		-1.236	1.6648	0.1159
Balanced	**Est**	36	-4E-04	0.0004	0
	t_stat		-0.79	2.0189	-0.008
USEquity	**Est**	55	-0.002	0.0009	-7E-04
	t_stat		-1.583	1.3533	-0.954

Table 3

Selectivity and Components of Timing Ability in a Three-Factor Model

This table reports estimate of alpha and its decomposition into market timing, volatility timing, and selectivity. The FF 3 factors define the benchmark. The sample covers January of 1984 through December of 2010. The GMM with a Newey-West covariance matrix with three lags is used for estimation. Level timing is the estimate of α_m, volatility timing is the estimate of α_σ and selectivity is the estimate of $\alpha_{s.}$ Total fund alpha is the sum of the components. NObs is the number of the time series observations of the fund group used for the estimation. Panel A reports the GMM estimates of a and b.

Panel A:
Estimates of the Market-wide Parameters:

		Nobs	a	b1(mkt)	b2(smb)	b3(hml)
	est	108	1.05	2.73	-0.29	1.85
	t_stat		18.59	1.42	-0.11	1.11

Panel B:
Performance

			Level Timing	**Volatility Timing**	**Combined Timing**	**Selectivity**	**Total Alpha**
AssetAllocation	**est**	71	-0.001	0.0003	-6E-04	0.0007	0.0001
	t_stat		-0.804	0.663	-0.642	0.8703	0.0734
Balanced	**est**	79	0.0006	0.0002	0.0008	-0.001	-2E-04
	t_stat		0.7611	0.8319	1.229	-1.646	-0.306
USEquity	**est**	103	-2E-04	0	-2E-04	0.0014	0.0011
	t_stat		-0.152	-0.091	-0.227	1.0886	0.7523

Table 4

Selectivity and Components of Timing Ability in a Four-Factor Model

This table reports estimate of alpha and its decomposition into market timing, volatility timing, and selectivity. The Carhart 4 factors are the benchmarks. The sample covers January of 1984 through December of 2010. The GMM with a Newey-West covariance matrix with three lags is used for estimation. Level timing is the estimate of α_m, volatility timing is the estimate of α_σ and selectivity is the estimate of α_s. Total fund alpha is the sum of the components. NObs is the number of the time series observations of the fund group used for the estimation. Panel A reports the GMM estimates of a and b.

Panel A:
Estimates of the Market-wide Parameters:

		Nobs	a	b1(mkt)	b2(smb)	b3(hml)	b4(umd)
	est	108	1.26	4.27	3.15	7.26	7.40
	t_stat		8.21	1.87	1.11	2.75	2.7

Panel B:
Performance

			Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha
AssetAllocation	**est**	71	-9E-04	0.0001	-8E-04	0.0007	0
	t_stat		-0.635	0.1534	-0.72	0.8294	-0.03
Balanced	**est**	79	0.0003	0	0.0003	-6E-04	-2E-04
	t_stat		0.3158	0.0631	0.4733	-0.844	-0.286
USEquity	**est**	103	0.0004	-2E-04	0.0002	0.0006	0.0008
	t_stat		0.2688	-0.21	0.2174	0.4008	0.5117

Table 5

Market Timing in a Conditional CAPM

This table reports estimates of market level timing, volatility timing ability and their sum. The conditional model assumes that the market wide coefficients in the stochastic discount factor are linear functions of the lagged instruments, Z. The sample period is January of 1984 through December of 2010 and quarterly data are used. The coefficients aj and bj, j=0,…,4 are reported in the following order: Intercept, dividend yield, Tbill yield, TERM and DEF. The GMM with a Newey-West covariance matrix with three lags is used in estimation.

Panel A:
Estimates of the Market-wide Parameters:

		NObs	a0	a1	a2	a3	a4
	est	108	1.136	8.62	-2.38	0.10	-5.03
	t_stat		3.3	0.62	-0.39	0.01	-0.34
			b0	**b1**	**b2**	**b3**	**b4**
	est		16.90	1006.45	-333.21	-273.78	-1298.16
	t_stat		2.12	3.64	-2.81	-1.77	-3.21

Panel B:
Performance

			Level Timing	Volatility Timing	Sum
AssetAllocation	**est**	32	0.0001	0.0002	0.0003
	t_stat		0.0982	0.3679	1.3649
Balanced	**est**	35	-2E-04	0.0002	-1E-04
	t_stat		-0.57	0.355	-0.612
USEquity	**est**	54	-1E-04	0.0003	0.0002
	t_stat		-0.19	0.8614	1.1805

Table 6

Components of Performance in a Conditional Three-factor Model

This table reports estimates of market level timing, volatility timing ability, selectivity and their sum. The reduced-form model assumes that the market wide coefficients in the stochastic discount factor are linear functions of the lagged instruments, Z. The Fama-French FF 3 factors are the benchmarks. The sample period is January of 1984 through December of 2010 and quarterly data are used. The coefficients aj and bj, j=0,…,4 are reported in the following order: Intercept, dividend yield, Tbill yield, TERM and DEF. The GMM with a Newey-West covariance matrix with three lags is used in estimation.

Panel A:
Estimates of the Market-wide Parameters:

		Nobs	a0	a1	a2	a3	a4
	est	108	-0.13	-14.97	11.44	23.95	89.28
	t_stat		-0.18	-0.59	1.05	1.42	1.51
b(mkt)			**b0**	**b1**	**b2**	**b3**	**b4**
	est		17.38	1841.23	-416.64	-438.311	-2351.37
	t_stat		1.53	4.45	-2.46	-1.54	-3.34
b(smb)			**b0**	**b1**	**b2**	**b3**	**b4**
	est		0.97	-825.54	-290.55	406.70	2008.59
	t_stat		0.07	-1.84	-1.2	1.27	1.97
b(hml)			**b0**	**b1**	**b2**	**b3**	**b4**
	est		-25.26	549.62	236.41	315.51	-513.38
	t_stat		-1.47	1.17	1.12	0.81	-0.85

Panel B:
Performance

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Sum
AssetAllocation	**est**	71	0.0007	-6E-04	0.0002	0.0011	0.0013
	t_stat		0.3804	-0.381	0.1525	1.3238	1.2003
Balanced	**est**	79	0.0024	-0.002	0.0009	-6E-04	0.0003
	t_stat		1.3474	-1.07	1.5119	-1.006	0.4848
USEquity	**est**	103	-3E-04	0.0004	0.0001	0.0014	0.0015
	t_stat		-0.24	0.4032	0.1569	1.2613	1.0427

Table 7

Performance of individual funds, sorting on various predetermined fund characteristics.

This table reports estimate 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. 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
quintle 1 (lowest)	**est**	103	0	-7E-04	-7E-04	0.0016	0.0009	0.0013
	tstat		-0.02	-0.71	-0.88	1.51	0.77	2.22
quintle 2	**est**	103	0.0006	-2E-04	0.0003	0.0006	0.001	0.0014
	tstat		0.34	-0.20	0.35	0.61	0.66	1.90
quintle 3	**est**	103	0.0009	-7E-04	0.0002	0.0002	0.0004	0.0016
	tstat		0.53	-0.63	0.18	0.12	0.25	2.08
quintle 4	**est**	103	0.0002	-6E-04	-4E-04	0.0006	0.0002	0.0015
	tstat		0.13	-0.50	-0.24	0.34	0.16	1.69
quintle 5 (highest)	**est**	103	0.0006	-3E-04	0.0003	0.0001	0.0004	0.0025
	tstat		0.37	-0.34	0.24	0.06	0.28	1.96
quintle 5 - quintle 1	**est**	103	0.0006	0.0004	0.001	-0.002	-4E-04	0.0012
	tstat		0.44	0.55	0.88	-1.21	-0.43	1.14

Panel B **Sorting on R-squares**

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	DGTWcs
decile 1 (lowest)	**est**	46	0.0042	-0.001	0.0033	0.0055	0.0087	0.0032
	tstat		0.87	-0.23	1.10	1.51	1.97	1.39
decile 2	**est**	46	0.0051	-0.002	0.0031	0.0044	0.0074	0.0032
	tstat		0.80	-0.36	0.88	1.22	1.59	1.33
decile 3	**est**	46	0.0047	-0.001	0.0036	0.0036	0.0072	0.003
	tstat		0.82	-0.23	1.16	1.16	1.72	1.24
decile 4	**est**	46	0.0045	-0.002	0.0029	0.0025	0.0054	0.0014
	tstat		1.23	-0.52	1.35	0.91	1.56	0.69

decile 5	**est**	46	0.0033	-0.001	0.0022	0.0025	0.0047	0.0026
	tstat		1.45	-0.77	1.27	1.12	1.54	1.19
decile 6	**est**	46	-2E-04	0	-2E-04	0.0016	0.0014	0.0021
	tstat		-0.14	0.03	-0.17	1.17	0.82	0.90
decile 7	**est**	46	-0.001	0.0014	0.0001	0.0016	0.0017	0.0019
	tstat		-0.47	0.76	0.06	1.45	1.11	0.97
decile 8	**est**	46	-0.001	0.0017	0.0006	0.0007	0.0013	0.0014
	tstat		-0.52	1.11	0.39	0.58	0.90	0.80
decile 9	**est**	46	-7E-04	0.0012	0.0005	0.0001	0.0006	0.0012
	tstat		-0.25	0.71	0.19	0.07	0.44	0.76
decile 10 (highest)	**est**	46	-3E-04	0.0004	0	-5E-04	-5E-04	0.0002
	tstat		-0.10	0.18	0.01	-0.23	-0.34	0.23
decile 10-decile1	**est**	46	-0.005	0.0014	-0.003	-0.006	-0.009	-0.003
	tstat		-0.71	0.33	-0.67	-1.23	-2.06	-1.28

Panel C **Sorting on Active Shares**

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	DGTWcs
quintle 1 (lowest)	**est**	68	0.0039	-0.002	0.0015	0.0004	0.0018	0.0014
	tstat		1.79	-1.49	1.33	0.29	1.77	1.87
quintle 2	**est**	68	0.0023	-0.003	-4E-04	0.0008	0.0004	0.0026
	tstat		1.10	-1.68	-0.43	0.59	0.35	1.75
quintle 3	**est**	68	-7E-04	-7E-04	-0.001	0.001	-4E-04	0.0023
	tstat		-0.20	-0.22	-0.84	0.61	-0.21	1.54
quintle 4	**est**	68	0.0001	-9E-04	-8E-04	-8E-04	-0.002	0.0022
	tstat		0.02	-0.32	-0.36	-0.37	-0.69	1.10
quintle 5 (highest)	**est**	68	-0.001	-9E-04	-0.002	0.0001	-0.002	0.0029
	tstat		-0.25	-0.25	-0.83	0.06	-0.58	2.02
quintle 5 - quintle 1	**est**	68	-0.005	0.0016	-0.004	-2E-04	-0.004	0.0014
	tstat		-1.19	0.45	-1.39	-0.09	-1.22	1.06

Table 8:

Performance of individual funds, sorting on various predetermined fund characteristics.

This table replicates Table 7, restricting the analysis to (1984:01-1999:12). This table reports estimate of alpha and its decomposition into market timing, volatility timing, and selectivity. The Carhart 4 factors are the benchmarks. The sample covers January of 1984 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).

Sorting on Expense ratios

Panel A

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	DGT Wcs
decile 1 (lowest)	**est**	60	-5E-04	-9E-04	-0.001	0.0022	0.0008	0.0022
	tstat		-0.27	-1.119	-0.991	2.0342	0.7544	3.4799
decile 2	**est**	60	-6E-04	-8E-04	-0.001	0.0019	0.0005	0.0006
	tstat		-0.262	-0.596	-0.767	1.1491	0.4761	0.7944
decile 3	**est**	60	-0.001	0.0011	-3E-04	0.0009	0.0005	0.0018
	tstat		-0.658	0.6742	-0.232	0.6535	0.3387	1.9482
decile 4	**est**	60	-6E-04	0.0005	-1E-04	0.0018	0.0017	0.0012
	tstat		-0.289	0.3604	-0.078	1.4021	1.3527	1.1475
decile 5	**est**	60	-9E-04	-4E-04	-0.001	0.0012	-1E-04	0.0021
	tstat		-0.425	-0.289	-0.71	0.8141	-0.066	1.928
decile 6	**est**	60	-0.003	0.0001	-0.003	0.0021	-6E-04	0.0017
	tstat		-1.082	0.0713	-1.776	1.1359	-0.319	1.9687
decile 7	**est**	60	-0.003	-0.001	-0.004	0.0026	-9E-04	0.0026
	tstat		-0.994	-0.495	-1.638	0.9709	-0.565	2.7159
decile 8	**est**	60	-0.004	0.0005	-0.003	0.0024	-8E-04	0.0014
	tstat		-1.936	0.3323	-1.874	1.0917	-0.585	0.9808
decile 9	**est**	60	-0.001	-0.001	-0.002	0.0023	0.0004	0.0039
	tstat		-0.503	-0.74	-1.194	1.2596	0.2316	2.6309
decile 10 (highest)	**est**	60	-0.002	0.0009	-7E-04	0.0027	0.002	0.0041
	tstat		-0.785	0.9286	-0.479	1.5912	1.3438	2.274
decile 10-decile1	**est**	60	-0.001	0.0019	0.0007	0.0004	0.0012	0.0019
	tstat		-0.431	1.5264	0.423	0.3091	0.7295	1.2385

Panel B

Sorting on Active Share

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	DGTWcs
decile 1 (lowest)	**est**	39	0.0038	-0.003	0.0006	0.0006	0.0012	0.0021
	tstat		1.6271	-1.173	0.4247	0.3253	1.1277	2.4054
decile 2	**est**	39	0.0015	-0.001	0.0004	-6E-04	-2E-04	0.0024
	tstat		0.7228	-0.715	0.2495	-0.294	-0.132	1.7585
decile 3	**est**	39	0.0005	-7E-04	-3E-04	0.0004	0.0001	0.0035
	tstat		0.1775	-0.396	-0.183	0.2276	0.0811	2.1718
decile 4	**est**	39	0.0028	-0.003	-3E-04	0.0003	0	0.0041
	tstat		0.9732	-1.014	-0.165	0.125	-0.003	1.7418
decile 5	**est**	39	-8E-04	0.0003	-5E-04	-3E-04	-8E-04	0.0032
	tstat		-0.207	0.1195	-0.238	-0.154	-0.472	1.8402
decile 6	**est**	39	-0.007	0.0059	-7E-04	-0.002	-0.003	0.0018
	tstat		-1.044	1.0335	-0.217	-0.936	-1.992	0.9458
decile 7	**est**	39	0.0017	0.0019	0.0035	-0.005	-0.002	0.0043
	tstat		0.2874	0.393	1.0446	-1.483	-0.581	1.7256
decile 8	**est**	39	-0.005	0.0025	-0.002	-0.003	-0.006	0.0036
	tstat		-0.806	0.4564	-0.621	-1.1	-2.24	1.3062
decile 9	**est**	39	-0.003	0.0029	0	-0.004	-0.004	0.0044
	tstat		-0.494	0.5779	0.0011	-1.326	-1.629	2.1505
decile 10 (highest)	**est**	39	-0.009	0.0062	-0.002	0.0007	-0.002	0.0021
	tstat		-1.272	0.9216	-0.759	0.3323	-0.549	1.0252
decile 10-decile1	**est**	39	-0.012	0.0093	-0.003	0.0001	-0.003	0
	tstat		-1.456	1.0524	-0.738	0.024	-0.95	-0.003

Panel C

Sorting on Return Gap

		Nobs	Level Timing	Volatility Timing	Combined Timing	Selectivity	Total Alpha	DGTWcs
decile 1 (lowest)	**est**	59	0.0041	0.001	0.0051	0.0002	0.0053	0.0027
	tstat		1.3988	0.5575	1.8776	0.145	2.0506	1.4211
decile 2	**est**	59	-4E-04	0.0013	0.0009	0.0014	0.0024	0.0025
	tstat		-0.144	0.8075	0.3322	1.2027	0.8383	1.9754
decile 3	**est**	59	-0.004	-3E-04	-0.004	0.0036	-4E-04	0.0017
	tstat		-1.769	-0.209	-2.418	2.229	-0.342	1.8972

decile 4	**est**	59	0.0009	-8E-04	0	0.0011	0.0012	0.0015
	tstat		0.5346	-0.702	0.0356	0.7756	0.9781	1.7978
decile 5	**est**	59	-0.002	-0.001	-0.003	0.002	-0.001	0.0021
	tstat		-0.98	-1.044	-2.357	1.5405	-0.787	2.1174
decile 6	**est**	59	-0.003	0.0011	-0.002	0.0025	0.0004	0.0013
	tstat		-1.369	0.5171	-1.227	1.4787	0.2968	1.3189
decile 7	**est**	59	-0.004	0.0002	-0.004	0.0023	-0.001	0.0026
	tstat		-1.318	0.083	-1.702	1.4458	-0.813	3.1814
decile 8	**est**	59	-0.001	-0.001	-0.003	0.0023	-4E-04	0.0021
	tstat		-0.618	-0.643	-1.409	1.6609	-0.217	2.1
decile 9	**est**	59	-0.003	-0.001	-0.005	0.0034	-0.001	0.0029
	tstat		-1.466	-0.628	-2.688	1.8845	-0.593	2.494
decile 10 (highest)	**est**	59	-0.003	0.0008	-0.002	0.0011	-0.001	0.003
	tstat		-0.922	0.4347	-0.742	0.4289	-0.331	1.2691
decile 10-decile1	**est**	59	-0.007	-1E-04	-0.007	0.0009	-0.006	0.0002
	tstat		-1.829	-0.055	-2.156	0.3793	-1.42	0.1184

Panel D **Sorting on Turnover as in Wermers(2000)**

		Nobs	**Level Timing**	**Volatility Timing**	**Combined Timing**	**Selectivity**	**Total Alpha**	**DGTWcs**
quintle 1 (lowest)	**est**	52	-0.005	0.002	-0.003	0.0021	-5E-04	-0.001
	t_stat		-1.489	0.8043	-1.114	1.6113	-0.302	-0.506
quintle 2	**est**	52	-0.001	-0.001	-0.002	0.0017	-4E-04	0.0009
	t_stat		-0.396	-0.589	-1.34	0.864	-0.26	0.3603
quintle 3	**est**	52	-0.007	0.0032	-0.004	0.0017	-0.002	-6E-04
	t_stat		-1.727	1.0113	-1.25	0.8779	-0.654	-0.286
quintle 4	**est**	52	-0.003	-0.004	-0.006	0.0039	-0.003	0.0031
	t_stat		-1.011	-1.364	-3.694	1.9764	-1.202	2.7353
quintle 5 (highest)	**est**	52	-0.002	-5E-04	-0.002	0.0037	0.0015	0.006
	t_stat		-0.637	-0.22	-1.165	1.3585	0.5711	2.1898
quintle 5 - quintle 1	**est**	52	0.0029	-0.002	0.0005	0.0015	0.002	0.0071
	t_stat		0.9126	-0.824	0.1694	0.5928	0.5729	1.9623