Improved Forecasting of Mutual Fund Alphas and Betas *

Harry Mamaysky †
Morgan Stanley
1585 Broadway
New York, NY 10036

Matthew Spiegel ‡
Yale School of Management
P.O. Box 208200
New Haven, CT 06520-8200

Hong Zhang §
INSEAD
1 Ayer Rajah Avenue
Singapore 138676

February 20, 2006

*We thank David Musto whose critique of an earlier paper lead to the creation of the eight factor model used in this one. We also thank Mark Carhart comments that lead us to explore the impact of potential sample biases in the CRSP mutual fund database. Additional thanks go to Jonathan Berk, Joshua Coval, Wayne Ferson, William Goetzmann, and Peter Starr for their comments. Finally, we thank seminar participants at INSEAD, Rutgers, the University of Michigan at Ann Arbor, the University of Calgary, and the University of Alberta and conference participants at the 2005 Winter Finance Summit, the 2005 Meetings of the Western Finance Association, and the 2005 Meetings of the European Finance Association.

†Email: harry.mamaysky@morganstanley.com, phone (212) 761-1022, fax (212) 761-9589.

‡Email: matthew.spiegel@yale.edu, phone (203) 432-6017, fax (203) 436-0630, som.yale.edu/∼spiegel.

§Email: hong.zhang@insead.edu, phone 65-67995340, fax 65-67995366.

This is a pre-copyedited, author-produced PDF of an article accepted for publication in the Review of Finance following peer review. The version of record [insert complete citation information here] is available online at: < http://dx.doi.org/10.1093/rof/rfm018 >.

Improved Forecasting of Mutual Fund Alphas and Betas

Abstract

This paper proposes a simple back testing procedure that is shown to dramatically improve a panel data model's ability to produce out of sample forecasts. Here the procedure is used to forecast mutual fund alphas. Using monthly data with an OLS model it has been difficult to consistently predict which portfolio managers will produce above market returns for their investors. This paper provides empirical evidence that sorting on the estimated alphas populates the top and bottom deciles not with the best and worst funds, but with those having the greatest estimation error. This problem can be attenuated by back testing the statistical model fund by fund. The back test used here requires a statistical model to exhibit some past predictive success for a particular fund before it is allowed to make predictions about that fund in the current period. Another estimation problem concerns the use of a single statistical model for all available mutual funds. Since mutual funds often, but not always, employ dynamic trading strategies their betas move over time in a ways that differ from fund to fund. Since no one statistical model is likely to fit every fund, the result is a great deal of misspecification error. This paper shows that the combined use of an OLS and Kalman filter model increases the number of funds with predictable out of sample alphas by about 60%. Overall, a strategy that uses very modest ex-ante filters to eliminate funds whose parameters likely derive primarily from estimation errors produces an out of sample risk adjusted return of over 4% per annum.

JEL Classification: G12, G13.

Over the last twenty years the mutual fund industry has grown at an incredible rate, and this has naturally attracted a lot of attention from the academic and financial community. One area of particular interest has been whether or not it is possible to identify fund mangers with skills that investors can capitalize on. The approach taken in this literature has been to apply a single statistical model to every available fund. But at any one time funds exhibit substantial differences in their strategies and holdings (Brown and Goetzmann (1997)). This makes it likely that any one statistical model will be incorrectly specified for at least some funds in the data pool. Also, as noted by Timmermann and Granger (2004) behavioral changes over time can cause a model that fits a subject in one time interval to fail in another. As a result many of the estimated parameters values used to forecast fund returns (especially those in the extreme deciles) may reflect misspecification error rather than reality. This paper proposes a very simple back testing procedure to help alleviate this problem. The results indicate that with back testing even simple OLS models, which have previously been found to exhibit little predictive power, produce useful forecasts for large subsets of funds. Back testing is also shown to generate substantial improvements for the non-linear models tested here as well. Overall back testing along with a combination of models can produce portfolios encompassing over 15% of the mutual fund population that yield economically and statistically significant predictable above market performance in any one period.

In general, the problem with using a single model for every fund in a time series panel database is that sorting on the estimated alphas (or any other attribute) may simultaneously sort on misspecification errors. If this happens, then models may not select funds with predictable superior performance as "best" but rather those with the poorest parameter estimates as these will tend to be the most extreme.[1] One possible way to help identify possible misspecification errors, and the one pursued here, is the following algorithm: prior to using a model to forecast a particular subject's performance in the current period it must first generate acceptable out of sample forecasts in the recent past. This helps to avoid using the model's in sample attributes to determine if it will do well out of sample. Intuitively, if one wishes to use a model to find funds that will produce future above (or below) market returns it seems natural to require at least some past success in this regard. If an extreme high alpha fund does under perform in the recent past one might strongly suspect that there may be some problem with the model's forecasting ability, at least for the moment. The same should be true for an extreme low alpha fund that over performs. Consequently, the back test proposed here requires a model to correctly predict the sign of a fund's excess return in the previous month before allowing it to make predictions about the fund's future behavior. Applying even this simple criteria yields a dramatic improvement of the risk adjusted return for top mutual fund deciles selected by either the one- or Carhart's four-factor OLS model. Monthly returns jump from initial values of -8 basis points (bps) and

[1]This argument is similar to the size anomaly critique found in Berk (1995). There he argues that misestimated betas will lead to the appearance that small firms outperform large ones on a risk adjusted basis. Roughly, all else equal, higher discount rates lead to smaller market capitalizations. To the degree that a model misestimates beta the firm's market capitalization will then proxy for the true cash flow risk.

18bps to the economically and statistically significant values of 21bps and 37bps, respectively.

Beyond that this paper also shows the benefits of simultaneously using multiple models. Back testing implies that any one model will not be used to generate forecasts for every single subject. Thus, by using multiple models the set of funds for which one might produce useful forecasts potentially widens. To illustrate the point a dynamic Kalman filter model is tested along with the standard rolling OLS model.[2] These models provide a good pairing as they offer very different costs and benefits and (with the help of back testing to clean out misspecification errors) capture different types of managerial skills. Rolling OLS models are simple to estimate. But their validity requires that a portfolio's parameters drift slowly over time if at all. If a fund actively trades securities during the (typically) estimated five year window the resulting parameter estimates may not accurately represent the current situation.[3] A Kalman filter model can potentially adapt itself to such changes and avoid this problem.[4]

Similar to the OLS case, with back testing the Kalman filter models successfully select funds with out of sample risk adjusted abnormal returns in excess of 3.5% per annum. More importantly, each model selects a relatively unique set of funds with an overlap of only about a third. This implies that it is possible to find a remarkably wide variety of funds with positive (or negative) predictable risk adjusted returns if one is willing to employ a variety of models. Hence, instead of running a horse race among different models and picking a "winner," this paper demonstrates the benefit of simultaneously using more than one model. By going from one model to two the set of funds with predictable super normal returns increases by about 50%. Meanwhile, funds jointly selected by the OLS and Kalman models within the top decile, presumably the funds run by managers with more than one type of managerial talent, can have risk adjusted returns as high as 6.0% per annum.

Whether or not statistical models can identify fund managers that will produce positive risk adjusted returns for their investors has implications regarding the general functioning of markets. If such skills do not exist then this calls into question the value of fundamental analysis and active management, at least for mutual fund investors. On the other hand, if such skills do exist but are somehow bid away when discovered then this offers support for the Berk and Green (2004) hypothesis regarding fund returns in an efficient competitive environment.[5] Overall, this paper's findings offers support for at least part of their thesis;

[2]See Mamaysky, Spiegel, and Zhang (2005) for a detailed derivation of the Kalman model based on dynamic selection ability.

[3]One solution can be found in Grinblatt and Titman (1994). The methodology they use avoids a direct comparison against a specific portfolio, and instead uses an "endogenous" benchmark. However, their technique requires knowledge of the fund's actual composition, which may not always be available. Ferson and Khang (2002) extend the technique to condition the portfolio betas on exogenous variables such as macro economic data.

[4]Grinblatt and Titman (1989) also propose a technique that can detect market timing abilities that arise from a fund's dynamic asset allocation strategy and implement it in their 1994 paper. However, as Ferson and Schadt (1996) point out correlations between factor loadings and market returns may also be due to predictable changes in time varying expected returns, and thus implement a technique for handling this case.

[5]The published model excludes the cost of searching for funds with superior managers and thus states that

managerial skill exists but its benefit to mutual fund investors is short lived.

Since Carhart's (1987) paper on the relationship between mutual fund and momentum returns performance studies have moved onto the use of more comprehensive data sets and/or improved methodologies. One approach has been to use the underlying holdings data as in Cohen, Coval, and Pástor (2005), and Kacperczyk, Sialm, and Zheng (2005a). Another has been to use Bayesian models as in Avramov and Wermers (2005) and Busse and Irvine (2005) or daily data as in Bollen and Busse (2004). This paper can be seen as both complimenting and extending this literature. The simple back test proposed here allows even the one factor OLS model to reliably identify funds that will produce above market returns. This observation suggests that misspecification error is conceivably the main impediment preventing traditional models from identifying superior funds. It also suggests that it is perhaps too early to give up on traditional models and data sets (e.g. Hendricks, Patel, and Zechhauser (1993), and Brown and Goetzmann (1995)) in the quest to find managerial talent ex-ante.

The back testing procedure suggested here can also potentially strengthen the findings and methodologies within the above and other related papers. Indeed, this has started happening. Kacperczyk, Sialm, and Zheng (2005b) use a variant of the back test proposed here to improve their model's ability to forecast fund returns based on the difference between observed returns and those calculated from the reported holdings data.[6] Future research will undoubtedly show that other (better) back testing procedures can be employed. Still, it should be emphasized that the goal here is not to produce an "optimal" estimator or back test. Rather, the goal of this paper is to develop an effective and simple procedure that is likely to be robust across a variety of possible situations and thus potentially a heuristic for future research. Having said that, it is not obvious that more complex filters alone will in fact yield better results. The difficulty is that misspecification errors can influence the values generated by any statistical model in a variety of ways and thus to the degree more complex filters are less robust they may yield inferior results. In fact, this paper finds just that. When the estimated betas are used as part of the back test the resulting portfolios exhibit poorer performance. Alternatively, one might think to use a model's diagnostics like the R^2, or a particular t value. But, misspecification error may lead not only to large parameter estimates but also to erroneously good in sample diagnostics. The problem is that admitting that a model may be misspecified is tantamount to admitting that "we do not know what we do not know." This can be seen in the negative results generated by Bossaerts and Hillion (1999) and Goyal and Welch (2006). Both studies try various model switching criteria and find no out of sample predictability regarding the equity premium (Goyal and Welch even find it degrades the resulting portfolio's performance). Both studies also suggest that "model instability" may be to blame, which is a form of misspecification error. Thus, simpler (and potentially more robust) tests may in the end produce the best results.

Since mutual funds use "closing prices" to calculate end of day net asset values, any

investors literally earn zero excess returns from their mutual fund investments. However, adding in a cost for actually searching will (inside their framework) imply that well designed strategies should be profitable.

[6] They cite this paper as the source for the procedure.

model may appear to select funds with positive abnormal returns by taking advantage of the resulting stale pricing problem.[7] This is an important issue since such abnormal returns are not related to managerial ability but rather the use of a poor (and exploitable) pricing algorithm for calculating end of day net asset values. To control for this possibility the paper introduces a "fifth through eighth factor;" the four factor returns from month $t-1$ to month t. This "eight factor model" should capture any stale pricing impacts as returns due to past market returns should be absorbed by the historical return parameters. These stale pricing factors typically reduce the out of sample returns by about four to eight basis points per month but typically do not eliminate them.

Models seeking out managerial talent may also produce spurious statistics indicating their success due to the "omitted return" problem within the CRSP data base (Carhart (1995) and Elton, Gruber, and Blake (2001)). This problem arises because there are some deceased funds for which it has not been possible to recreate an entire series of returns prior to 1983. When that happens the CRSP files have a missing value codes for those months and as a result the funds are lost to the analysis conducted here. Because omitted returns are associated with funds that subsequently died they create problems similar to those associated with survivorship biases. To test whether or not this paper's results are likely to arise from the CRSP file's omitted return bias several tests are conducted. One set examines only those funds with omitted returns and estimates that their exclusion probably adds less than half a *basis point* to the paper's reported results. Another test using only post 1985 data, when well under half a percent of all funds suffer from the omitted return problem, finds that the results are qualitatively similar to those from the entire post 1970 data set. This further indicates that the CRSP file's omitted return bias is not driving the paper's results.

This paper is also related to the general econometric literature on forecasting. For example White (2000) shows how to produce exact p-values given a postulated distribution function for the hypothesis that some model is superior to the benchmark when several models are run on a data set. Pesaran and Timmermann (2005) examine model selection in real time and also seek tests to determine which among many models is the best at any moment in time. Here we are not interested in finding the best overall model, but rather seeing if a given model's performance can be improved by limiting its use to particular subjects within a data set. A more closely related article is Timmermann and Granger (2004). They discuss attempting to create forecasts under the assumption that markets if not instantly efficient will eventually work to invalidate any currently successful forecasting method. As they note in this environment forecasters will frequently need to change the model they use. It is worth

[7]The closing prices used to set mutual fund net asset values are in reality the last price a security traded at prior to 4:00PM eastern time in the U.S. For infrequently traded securities this price may have been recorded hours if not days earlier. Investors can take advantage of this by purchasing funds with infrequently traded securities after the broad market has gone up, and by selling these funds following a market decline. This strategy sometimes goes by the name of market timing and the academic literature on this issue has grown extensively in recent years. Papers by Chalmers, Edelen, and Kadlec (2001), Greene and Hodges (2002), Zitzewitz (2003), and Goetzmann, Ivkovic, and Rouwenhorst (2001) all show how investors can exploit stale net asset value prices in a variety of funds.

noting that the back testing procedure suggested here allows for potentially rapid switching across models to capture the changing behavior of particular subjects.

All of the empirical tests in this paper are conducted with data from the CRSP mutual fund files. Only equity funds (defined as funds with objective codes of AG, BL, GI, IN, LG, PM, SF, or UT) are used. This data is then augmented with the monthly CRSP value weighted return market index, and the Fama-French and Carhart factors as provided by the WRDS web site. While funds with load fees are included in the analysis (to be consistent with past studies) the results are essentially unchanged when they are excluded.[8]

The remainder of the paper proceeds as follows. Section 1 provides evidence regarding the impact of estimation errors on sorts. Section 2 examines the out of sample returns for the various models. Section 3 explores other back tests. Section 4 looks at the degree to which the models pick similar funds. Section 5 investigates the omitted return problem and compares the results in this paper to those in Carhart (1997). Section 6 concludes. The Appendix (Section 7) derives a dynamic model of mutual fund returns (used as an alternative to the typically employed static OLS specification) and explains how to estimate it via a modified Kalman Filter.

1 Alpha-Beta Relationships in the OLS Model

As noted above OLS models seeking to predict mutual fund returns typically begin by estimating a factor model within a five year rolling window. Next, funds are sorted on alpha and ranked by decile. Tests are then conducted on each decile's out of sample performance. However, sorting in this manner ranks funds both by their future alphas and the estimation error associated with each fund. In a factor model this is especially problematic since misestimated betas will themselves induce misestimated alphas. To see why consider a one factor model and imagine that the estimated beta is too low. If the fund actually holds a portfolio of stocks then in a typical year it will have a higher return than the risk free asset. With an underestimated value of beta the model will then try to fit the returns by raising alpha. The opposite will also hold; overestimated betas will typically lead to underestimated alphas.

Table 1 examines how estimation errors feedback between alphas and betas. For each fund a one or four factor model of the form

$$r_{it} - r_t = \alpha_t + \beta_i{}'(r_{mt} - r_t) + \epsilon_t \tag{1}$$

is estimated. Here r_{it} is fund i's period t return, r_t the risk free rate, r_{mt} the vector of factor returns, and ϵ_t an error term. The estimated parameters are α and the vector β. In an ideal world (without estimation error) sorting on alpha or beta should not lead to

[8]Readers can, to some degree, easily verify this for themselves. Out of sample statistics are provided for the top 5, 10, and 20 funds as well as the top decile. All four groups generate similar results. Thus removing the funds with loads from the top decile leaves more than enough funds to populate a top 5, 10, and 20 grouping.

systematic patterns in the other parameter. However, as the table shows with real data it does. Column one displays the average estimated market β sorted by decile for the one factor CAPM model. Column two displays the average associated α for funds in that particular beta sorted decile. The first thing to note is the wide range of beta estimates. If one believes the OLS estimates then ten percent of equity funds have betas around .4 and another ten percent around 1.48. More striking are the alphas. There is a strong negative association between a fund's beta and its estimated alpha. The Spearman rank correlation coefficient is -.92 and statistically significant at any reasonable level.

The four factor model delivers results just as strong as the one factor model. Column three in Table 1 sorts on the four factor model's "return weighted beta" for each fund which is defined as

$$\beta_i' = \frac{\beta_{i1}\bar{r}_1 + \beta_{i2}\bar{r}_2 + \beta_{i3}\bar{r}_3 + \beta_{i4}\bar{r}_4}{\bar{\beta}_1\bar{r}_1 + \bar{\beta}_2\bar{r}_2 + \bar{\beta}_3\bar{r}_3 + \bar{\beta}_4\bar{r}_4}. \quad (2)$$

The $\bar{r}$ terms are the average returns for each factor, the $\bar{\beta}$ terms are the average estimated factor betas across funds, and the β_{ij} terms represent the estimate of factor j's loading for fund i. This metric is designed to indicate how sensitive a fund is to overall movements in the four factors. To make the metric's values correspond to those derived with the one factor model the denominator normalizes the average fund's return weighted beta to one. Thus, no matter how many factors one employs (one or more) the average value of the metric remains the same; one. Column four in the table lists the average alpha value for the funds in the corresponding return weighted beta deciles. As with the single factor model there is a nearly perfect negative correlation between the beta value and alpha values in each decile.

Table 1's alpha sorts (columns 5 through 10) show that the estimated alphas range over an implausible set of values. The lowest decile is predicted to lose a risk adjusted return of 10% over the next year, and the top decile to earn approximately 8.5% on a risk adjusted basis. Given how difficult it has been to find *any* fund group that constantly outperforms the market it is difficult to assign much credibility to any of the parameter estimates in either the top or bottom 3 deciles (60% of all the estimated parameters)! Columns 7 and 10 contain further evidence that the extreme alpha deciles contain unreliable parameter estimates. These columns display the average standard errors associated with each alpha decile. Simple observation yields a distinct U-shaped pattern with relatively large values (i.e. poor estimates) associated with the high and low deciles.[9] The only good news for the potential usefulness of the estimated parameter sorts appears in columns 6 and 9. These columns show that the estimated betas do not trend either up or down as one goes from one alpha decile to the next as one would expect under the null hypothesis.

Estimation error not only produces extreme parameter estimates but also unreliable ones as documented in Table 2. The out of sample tests displayed there derive from a three step procedure. First, either the one or four factor model is estimated using five years of data. Second, the following year's returns are regressed on the corresponding one or four factor

[9]We thank Wayne Ferson for suggesting this test.

model and the out of sample betas calculated. Third, the error between the predicted market beta and the actual market beta is calculated. These beta forecast errors are then pooled into deciles sorted by the predicted alphas. Under the null hypothesis the beta prediction errors should not be related to the predicted alphas. Yet, as the column labeled c_{MKT} (which is the average difference of the predicted out-of-sample betas minus the realized in-sample betas) shows this is not the case. Consider the results from the single factor model. Funds with low alpha values over predict future betas by .05, which is significantly different from zero at any reasonable level. At the other end the betas for high alpha funds tend to be too low by about .039 which is again statistically significant. Furthermore, there is a one to one relationship between the alpha decile and the size of the average forecast error; higher alphas lead to betas that are ever lower than reality. Again this is evidence regarding the feedback between the estimated alphas and betas. To at least some degree high alphas are due to underestimated betas and low alphas to overestimated betas.

The last column in Table 2 Panel A reports the standard deviation of the market beta forecast error by decile. As one can see there is a distinct U-shaped pattern with the highest values associated with the extreme deciles. This indicates that not only are the betas associated with the extreme alpha deciles biased but that the model has a difficult time making reliable out of sample forecasts.

Panel B from Table 2 repeats the analysis using the four factor model. Here there is no discernable trend in the market beta's errors across alpha deciles; instead there is a systematic bias. The uniformly negative values in column two indicate that the four factor model overestimates the average fund's market beta in every single decile. An examination of the last four columns shows that the four factor model suffers from the same U-shaped error pattern as the one factor model, but now across all four factors. It appears that the extreme alpha funds are those for which the model produces the most unreliable out of sample factor loadings.

2 Out of Sample Returns with Filtering

When forecasting mutual fund alphas the standard approach uses a single model across the entire panel data set. Table 3 displays the results from this exercise with monthly rebalancing, and serves as a benchmark from which any modifications to the statistical procedure can be judged. Panel A displays the four factor alphas based on the out of sample decile portfolios formed via each model. While the four factor OLS model does provide some predictive power (the top decile generates statistically significant returns) none of the other models do. Panel B extends the four factor model to eight (each factor's current and lagged value) to account for the stale pricing problem associated with mutual fund returns. This has a somewhat erratic impact on the four factor OLS model's performance. Portfolios using the top 5 and 10 funds no longer produce statistically significant out of sample alphas. On the other hand, the broader top 20 and top decile portfolios do. It is perhaps possible that the very highest alphas are arising from model's ability to exploit stale prices while the alphas for funds a bit further down the list are due to actual talent. In any event, even the best portfolios yield rather modest t-statistics making it difficult to assuredly reject the null.

Based on the results in Section 1 the negative results in Table 3 are likely due to the fact that the extreme decile alphas and betas are of suspect quality. Thus, intuitively, one might suppose that this problem can be alleviated by simply dropping funds if their estimated parameters lie outside of some region. Table 4 displays the impact of this filtering device on the portfolio returns. Comparing the results to those in Table 3 Panel A, one can see that the OLS models exhibit some improved out of sample performance when the beta range is substantially restricted, but none from restricting the alphas. For the dynamic Kalman filter model (see the Appendix for its derivation) both the alpha and beta restrictions provide some modest improvements. Overall though, simply dropping funds for which a model's coefficients are unreasonable does not appear to produce substantially better performing portfolios. However, many of the subsequent tests still drop funds with monthly alphas predicted to exceed 2% in absolute value. As will be seen this primarily provides a formal way to define, in a manner repeatable by others, when the Kalman filter optimization algorithm does not converge.

Alpha and beta filters are an indirect way of checking to see if a particular statistical model fits a particular fund. A more direct approach (and one that can be used either in combination with the alpha and beta restrictions or on its own) is back testing. The general idea is that before one uses a statistical model to make predictions, there should exist some evidence that it was successful in this regard some time in the past. Along these lines it seems reasonable to require predictive success in the period prior to the model's use as a forecasting device. The procedure works as follows: First, the model is estimated with sixty months of data up to time $t-2$ and a predicted alpha and a set of betas for $t-1$ are calculated. If the fund's realized above market excess return in period $t-1$ has the same sign as the predicted alpha, then the fund is added to the active pool. Otherwise the fund goes into the inactive pool. For the back test period this is equivalent to calculating the realized alpha by setting the market beta to one and all other betas, if any, to zero and thus does not depend upon the model's own estimated factor loadings. Second, for funds in the active pool the model is estimated using data through time $t-1$. If the estimated market beta lies between zero and two and the alpha between -.02 and +.02 the fund remains in the active pool. Otherwise it goes into the inactive pool. Third, the funds within the active pool are sorted by alpha and the decile and top fund portfolios are then constructed.

The above procedure does *not* introduce any selection biases. Any investor can use the three steps in real time to select funds. To further ensure that there are no selection biases the out of sample alphas are calculated using the standard OLS procedure on the realized portfolio returns. That is, portfolios are formed each period and the return to the portfolio calculated. The resulting return sequence is then regressed on the appropriate factor model (four or eight) and the estimated alpha is then reported. Note that the in sample estimated alphas and betas are not used at all when calculating the out of sample test statistics. Further, the standard OLS testing procedure is employed out of sample to further ensure that a particular model does not somehow feed itself risk estimates that will lead to a biased alpha value.

Note that the back testing procedure deliberately avoids using a model's estimated factor loadings to determine the accuracy of the model's alpha forecast in the prior month. As shown in Section 1 for the extreme alpha funds the estimated beta parameters are of questionable reliability. If these estimated (and likely biased) factor loadings are then used to calculate a fund's realized alpha in the test period the likely outcome will be to pass the model for use with that fund. Intuitively, it thus seems preferable to ignore the factor estimates when conducting the back test. For those interested in how more complex backtests may perform, later on the paper displays the results from using a model's own factor estimates in the back testing period. In short, for every model employed here this procedure leads to less reliable out of sample forecasts. Still, the top decile alpha sorted portfolios continue to produce positive and statistically significant out of sample returns for every model other than the single factor OLS.

Table 5 lists the results when both the backtest and the parameter filters are used to select funds for the active pool prior to creating portfolios based upon the sorted alphas. Unlike the results in Carhart (1997) and Avramov and Wermers (2005) the top OLS portfolios produce statistically significant positive returns. This is not only true of the top decile but also of the *ninth* decile. This implies that with some back testing it is possible to find a substantial number of funds that will, on average, produce positive predictable returns. The same findings hold for the Kalman filter model. If one further restricts attention to the top twenty or better funds the excess returns are about 4.4% per annum. For every model the rank correlation between the deciles and their returns exceeds .95. This is statistically significant at any reasonable level, providing additional evidence that the alpha sorts are picking up relative out of sample fund performance.[10] As a further test of the efficacy of the back testing procedure, the analysis in Table 5 was repeated but this time with the funds from the inactive pool. To save space the table is not reported here. But the resulting decile portfolios have no predictive power further verifying that back testing helps remove from a model's forecast those funds it is unable to accurately track.[11]

The fact that back testing dramatically improves all four models tested here indicates that the results are not due simply to data snooping. By contrast, data snooping typically yields a small subset of "successful" models from the larger group under consideration. Furthermore the improvement seen with back testing takes place not only in the high and low forecasted

[10]Like nearly every other study regarding mutual fund performance the results in Table 5 indicate that finding funds with predictably negative alphas is not difficult. A better question (and one not addressed here) is given how easy they are to identify what allows them to remain in business? Berk and Xu (2004) attribute their survival to the fact that a substantial number of investors do not pull their money from funds even after they have done poorly. This allows these managers to continue operating despite having a poor past and predictably poor future set of returns.

[11]The authors thank Joshua Coval for suggesting that we run a test on funds from the inactive pool.

alpha deciles but overall as well. Absent back testing the decile portfolios created with the in sample sorted alphas yield statistically insignificant rank correlations across deciles out of sample. However, with back testing the correlation coefficients rise dramatically and are statistically significant at any reasonable level.

Related to the issue of estimation is whether or not adding factors to the model improves its out of sample performance. For the OLS model the four factor model does appear to offer superior out of sample performance relative to the one factor model. The difference generally comes to about 2% per annum. For the Kalman filter model however the results are reversed. Under nearly every sample test comparison performed here, decisions based upon an in sample one factor model outperform decisions based upon an in sample four factor model. This indicates that if researchers try other nonlinear models they may find that the simpler in sample versions perform better out of sample.

To determine the degree to which the performance results in Table 5 may be due to a model's ability to exploit stale pricing problems Table 6 repeats the analysis using the eight factor model. Most of the models see their alphas decline by about five basis points per month. Nevertheless, the top deciles continue to produce statistically significant out of sample returns for all but the one factor OLS model. In addition, the t-statistic for the ninth decile's alpha equals 2.15 for the funds selected by the four factor OLS model, and 1.93 for the funds selected by the four factor Kalman filter model. If one restricts attention to the top 20, 10, or 5 funds then the addition of the lagged factors reduce the estimated alphas by about eight basis points per month for the four factor OLS model, but only three basis points for the Kalman filter model. This provides some initial evidence that the two models are indeed picking up different funds and selecting on different skill sets. More importantly, however, the t-statistics for the portfolios containing the top 10 or 20 funds remain above three for both the four factor OLS and one factor Kalman filter model. Thus, even when one controls for stale pricing a statistical model combined with back testing can still find a substantial number of funds that will outperform the four or eight factor market benchmark over the next month.

Table 7 examines the performance of each model with back testing but without the alpha and beta restrictions. Without the alpha and beta restrictions the Kalman filter model loses its predictive power. This is not too surprising. For nonlinear models convergence of the parameter search algorithm is not guaranteed. Researchers frequently get around this problem by reporting only the results for which the search algorithm "converged." However, without a clearly defined rule for determining if convergence has occurred one cannot guarantee that a particular result can be reproduced by other researchers. The alpha and beta restrictions provide the necessary definition in a way that can be clearly specified in advance. This prevents the creation of portfolios filled with funds for which the search algorithm "got lost" and yielded very high alpha values. However, even for the OLS models the alpha and beta restrictions provide some additional value over and above the back testing requirement (in terms of an increased point estimate but not to the degree that there is a statistically significant difference).

3 Other Back Tests

The out of sample performance tables discussed so far use a back testing procedure that does not depend upon a model's factor estimates. Table 8 displays the results from using a back test that instead relies on them. In this case the back test adjusts a fund's previous period return by using the statistical model's out of sample predicted betas for that period. Thus, the model is estimated with data up until month $t-2$. The resulting betas are then used to risk adjust the fund's realized return in period $t-1$. If the realized risk adjusted return then matches in sign the model's alpha prediction the fund goes into the active pool for period t. Using this procedure, across every model, the top decile fund portfolios produce a point estimates lower than those in Table 5 (in which the active pools do not depend on a model's beta estimates). Furthermore, for all but the single factor Kalman model the ninth decile fund returns are no longer statistically different from zero. Looking at the top 20, 10, or 5 funds the models in general see a significant degradation in the predictive ability and for many of the portfolios the returns are no longer different from zero at the standard significance levels.

The above results reinforce those from Section 1. For high and low alpha funds the beta estimates are simply not very reliable. Using them to then "help" pick funds likely to outperform therefore seems unlikely to succeed. Not necessarily because such funds do not exist, but because the in sample alpha sorts lead to estimation error sorts on the predicted betas which then feed back into poor alpha forecasts. To some degree this is expected given that Bossaerts and Hillion (1999) and Goyal and Welch (2006) both find that model selection criteria based on in sample statistics not only fail to improve the out of sample portfolio returns but actually reduce them. While both studies concern time variation in the equity premium, the general problem of forecasting returns with a battery of possible models is similar to the issues explored here.

Table 9 uses the in sample diagnostics of t value to sort mutual funds. No back testing or other filters are imposed. The overall result is similar to that of the benchmark case in Table 3. The four-factor model is the only model that can select superior mutual funds to some degree. The magnitude of the performance for top decile fund is similar to Table 3, while the t ratio is usually higher. One problem with the in sample t value, however, is at the negative side. In both Panel A and B the performance of bad funds does not persist, which contradicts the finding of Carhart (1997) and could be due to misspecification errors. The one factor Kalman model does select good top 5 out of sample funds. Other models do not seem to generate reliable in sample t value to select funds. Overall the table confirms that selection schemes based on in sample diagnostics perform worse than the back test proposed by this paper.

4 Fund Picks Across Models

While it appears that appropriate back testing of the fund selection models leads to better alpha estimates, it is possible that both the Kalman filter and OLS models produce very similar rankings. If that is the case then clearly one should limit future work to the OLS model as it is far simpler to calculate. To address this possibility Table 10 examines the degree to which the models select the same funds. Panel A reports the "Common Ratio." To construct this number funds are first ranked by each model. Next the union and intersection of the funds picked by models i and j for a particular decile are collected. The reported figure is the intersection divided by the union. Thus, for example, consider the row for decile ten and the figure of .43. This number implies that of all the funds ranked in the top decile by either the one factor OLS and Kalman model 43% were selected by both models. Note that both models are particularly likely to agree on which funds are particularly good or bad. Even so, by combining models one can expand the set of funds for which one can potentially predict performance by at least 50% to 60%.

Table 10 Panels B and C look at the fraction of funds in the active pools (i.e. those funds that are accepted by the filters) based upon either one model (Panel B) or a pair of models (Panel C). Take for example the .051 figure in Panel A in the decile 10 row under Model 1. This implies that the one factor OLS model had 51% of all funds in its active pool. Which, of course, means 5.1% of all funds were both in the active pool and ranked by the model in the top decile. Panel C show what happens to the active ratio when more than one model is employed. In this case the decile numbers are no longer one-tenth of the fraction of funds in the active pool due to picks common across models. However, the important point is that by using multiple models the fraction of all funds that are in the top decile increases from about 5% to 7.9% or more. Thus, by using two models simultaneously and back testing one obtains a top decile that contains nearly as many funds as one would have had by simply ranking all funds with a single model without back testing. More importantly, however, without the back test the resulting decile portfolios provide much less evidence that managerial talent can be detected. When combined these results indicate that there may exist a wide array of funds that can predictably outperform the market if one is willing to use a variety of models to find them.

If the Kalman and OLS models focus on different aspects of managerial talent then funds selected by multiple models might perform better than those selected by just one model. Table 11 examines this question by looking at the returns from a portfolio that selects only among the funds picked by a pair of models. Thus, when looking at models i and j, a fund is only selected if it is placed in the same decile by both models. The returns produced by these funds are in fact somewhat higher than those from the population as a whole. The top decile portfolio produces an excess return of nearly 6% for three out of the four possible model pairs. The one exception occurs for those funds picked by the two single factor models. Nevertheless, even here the point estimate increases somewhat from the returns found in Table 5. Comparing the top performing model in Table 5 with the top performing model pair in Table 11 yields an increased point estimate of 12 basis points per month. For the combination of the one factor OLS with the four factor Kalman the increased return is so large that it is statistically different from the average return produced by the two models alone (t-statistic of 2.36). Overall, then it appears that top decile funds selected by more than one model tend to produce higher returns than funds selected by just one model. This is at least consistent with the hypothesis that various models pick up on different aspects of managerial ability.

Even though the number of funds selected by any pair of models is a rather modest fraction of each model's selections it is possible that all of the excess returns come from that group. If so then this would imply that only a very small number of fund managers can produce above market returns and that one can find them by seeing if they are selected by multiple models. To test this hypothesis Table 12 examines the compliment of the funds in Table 11. For a fund to be included in a Table 12 portfolio for each model pair it must be selected by one for inclusion in the top decile but not by the other. Given the previous results it is not surprising that the portfolio returns produced by this group are somewhat lower than those found in Table 5. Nevertheless, every model's top decile continues to produce statistically significant positive alpha returns. Also, every model other than the one factor OLS model does so for the ninth decile portfolio as well. This reinforces the earlier conclusion that no single model appropriately describes the statistical properties of every single fund. It also buttresses the paper's thesis that a significant number of fund managers can produce predictable above market returns, but only if one uses a variety of statistical models to locate them and also back tests each model for its use with each fund.

5 The Carhart Results and the Impact of Missing Fund Returns

5.1 Comparison with Carhart's Results

One goal of this paper is to examine the degree to which back testing can be used to find mutual funds with superior out of sample performance. Perhaps though the efficacy of the back test proposed here really derives from differences in the estimation procedure used here versus Carhart (1995) and (1997). In both Carhart (1995) and (1997) a fund is ranked so long as it has at two and a half years of available data at the time a decision has to be made. Thus, the portfolios based upon three and five year training periods also include funds with shorter training periods.[12] Carhart's rational for this decision is to reduce the impact of what Elton, Gruber, and Blake (2001) call omission bias. As described in Carhart (1995) Appendix A, for the period prior to June 1983 the mutual fund database was created via

[12] While Carhart's (1997) analysis reports the results from portfolios formed on the basis of a three year training period, his unpublished (1995) dissertation includes those from a five year period as well. Since his results are similar across the two training periods the discussion that follows should be taken to encompass both techniques.

the use of several printed sources.[13] For some funds for some months the printed sources are incomplete. However, if the fund survived until 1983 it was generally possible to backfill the data using the Investment Company Data, Inc. (ICDI) database. Thus, most of the funds with missing data are those that were founded and terminated prior to 1983. Elton, Gruber and Blake (2001) find that these omissions lead to results that are similar to what one would expect from a survivorship biased sample; that is calculated alphas tend to be biased upwards.

While reducing the data requirements for a fund's inclusion to only two and a half years reduces the impact of omission bias, it potentially increases the impact of misspecification error. As Section 1 shows many of the estimated fund parameters are of questionable reliability even when a five year training period is used. A shorter training period can potentially produce even less reliable estimates, especially in a multifactor model where several parameters must be accurately forecast. This can be seen in Table 13 which compares the out of sample returns for sorted alpha portfolios using three and five year training periods. Panel A shows the standard four factor risk adjusted returns (on the left) along with their t-statistics (on the right). As one can see when funds are sorted based upon the alphas from a three year training period the top decile alpha returns are quite modest and one cannot reject the null hypothesis that they are zero at any reasonable level. However, when a five year training period is used the results change dramatically. Now the top decile earns an abnormal return of .12% a month which is statistically significant at the 5% level. If instead the portfolio concentrates on the top funds (20, 10, or 5) the results improve even further.

As with earlier tests it is possible that the results in Table 13 Panel A arise from the model's ability to exploit the stale pricing phenomenon inherent in reported fund returns. To test this Panel B regresses the alpha sorted portfolio returns on the eight factor model (the four current and four past factors). As one can see the realized returns do in fact drop somewhat. Under the eight factor model none of the decile portfolios produce statistically significant out of sample returns. Among the top fund portfolios only the one using the best 20 funds retains its statistically significant alpha although the point estimate drops by 3 basis points per month. Thus, for the one year holding period used by Carhart the evidence also indicates that at least some of the improved performance is due to stale pricing rather than managerial ability. Overall, though there remains strong evidence that using a five year training period to estimate the model produces substantially better out of sample predictions than one obtains by using a three year training period.

Related to the above issues is whether or not a one year holding period is or is not reasonable if one wishes to detect managerial performance. For example, Berk and Green's (2004) model (modified to include transactions costs) predicts that some mutual funds should exhibit detectable superior returns but not for long. If they are right then a one year holding period is likely to include long stretches past the point where a fund manager's ability can manifest itself. In such an environment a potentially superior technique for finding

[13] After June 1983 most of the fund data comes from ICDI's electronic database.

managerial ability is to reduce the holding period to one month. Compare Table 13 with annual rebalancing to Panels A and B in Table 3 with monthly rebalancing. With monthly rebalancing the four factor OLS model produces higher risk adjusted returns under both the four and eight factor models. Furthermore, this time the eight factor returns are statistically significant for both the top decile and the Top 20 portfolios. This provides at least some evidence for Berk and Green's hypothesis and for the use of shorter rebalancing periods if one wishes to detect managerial ability.

5.2 Impact of Omitted Returns

Given the results in Section 1 and Table 13 the next question that arises is whether or not the superior results reported here are due to a reduction in misspecification error or omission bias? Clearly the potential impact of the omission bias depends on the fraction of funds for which this is a problem. Table 14 shows that after the initial three years of data used in this study the omission bias impacts a very small fraction of funds. Starting in 1973 the fraction of funds dropped due to missing data never exceeds 1.05% and after 1979 it never exceeds .50% of all funds. Given how few funds are dropped due to missing data they seem unlikely to have more than a minor impact on the results reported here. For them to have a major impact they would need to both produce systematically above average alpha estimates in the training periods prior to their demise and very large negative alpha's in the holding periods that follow.

In order to determine whether or not funds with missing data are in fact likely to produce high training period alphas and low holding period alphas Table 15 examines their returns. Panel A shows that these funds produce remarkably poor return of about 1% per month below the market index. However, they apparently have low factor loadings and thus their alphas are quite a bit better than one might otherwise expect at -25 bps under the CAPM and -5 bps under the four factor model. Thus, while these funds produce below average returns they are not so low as to be that influential. As a rough estimate of their likely impact imagine that they are randomly included in any decile portfolio and represent approximately 1% of the fund population overall. In this case they would likely reduce any decile's alpha by no more than .50 bps or .10 bps under either the one or four factor model, a reduction so small that it would appear to be both economically inconsequential and unlikely to significantly impact the reported t-statistics.[14]

Table 15 Panel B takes the above analysis a step further and examines the likelihood that the funds with omitted returns will populate the top forecasted alpha decile if they could be included in the general analysis. Columns two and four in the first row show

[14]Let x represent the alpha of the excluded funds (-25 bps for example), y the alpha of the included funds, and z the ratio of the alpha derived from a portfolio of all funds to that of just the included funds. Then the following formula describes their relationship $.01x + .99y = zy$ or $z = .01(x/y) + .99$. Thus, unless the alpha from the excluded funds is negative and of greater magnitude than that of the included funds its impact will be less than twice its value. In this case that bounds it at .50 bps.

that during the training period these funds have low CAPM alphas but surprisingly high four factor alphas. However, the "Pass" columns indicate that the back testing algorithm proposed in this paper eliminates most of the excluded funds with high training period alpha estimates. Thus, while these funds might exhibit some tendency to populate the top four-factor forecasted alpha decile they are not a problem for either OLS model if it is first back tested fund by fund. The table's fourth row indicates that the holding period alphas for these funds are very close to zero. This provides further evidence that they are unlikely to significantly impact any reported holding period returns. Finally, the "AlphaDiff" row shows that the training period and holding period four factor alphas for those funds that pass the back test are remarkably similar. This further indicates that poor performing funds in this group would not be included in the top deciles. Of some interest however, is the fact that the full population of funds with omitted returns yield significantly lower holding period alphas relative to what they produced in the training period (-13 bps). This might help to explain why Carhart (1995) found that a three and five year training period produced similar holding period alphas under the four factor model.[15]

As a final check to ensure the results reported here are not unduly influenced by the omitted return bias Table 16 repeats the analysis from Table 5 Panel B but forms portfolios starting in 1985. As noted above the fraction of funds with missing data in the CRSP tapes drops to under .5% after 1979 and thus this sample period contains almost no omitted return biases. A comparison with Table 5 shows that the estimated alphas for the top funds are nearly unchanged by the later starting date. Naturally, the t-statistics are lower as the time frame is shorter but that is the only meaningful change. However, it is worth noting that even with the short time horizon the four factor OLS and one factor Kalman models continue to produce t-statistics above 1.95 for the top decile, top 20 and top 10 fund categories. Given these results and those in Section 5 it seems extremely unlikely that the results reported here are influenced to any great degree by the omitted return problem that arises from dropping funds with fewer than five years of reported returns in the CRSP database.

6 Conclusion

There is a long academic literature examining the question of whether or not it is possible to identify mutual fund managers who can produce predictable above market returns based only upon their past returns. However, these studies have used a single statistical model to describe every fund in their database. This is asking a lot of a single model. Fund managers

[15] Also recall that Carhart includes all funds with at least two and a half years of data. This implies that the three and five year training periods actually represent the longest training period used for each fund. For many funds the training period will be shorter. Furthermore, most of these funds will yield parameter estimates that are largely independent of the maximal training period used. (A fund with two and a half years of data will be completely unaffected by the maximal training period, while one with thirty one months of data will be nearly unaffected.) Since the same short training period funds appear in every sample with largely identical parameter estimates they may help explain why he finds that the training period makes little difference to the reliability of the forecasted alphas.

follow a wide variety of strategies (see for example Brown and Goetzmann (1997)) and it is unlikely that any one statistical model will accurately capture such a wide variety of dynamics. As a result, if the assumptions behind a model fit a particular fund's dynamics poorly then the resulting parameter estimates will suffer from misspecification error. In particular, a fund's alpha may be poorly estimated causing any sorts on this parameter to place it either at the top or bottom of any predicted performance list. Of course, going forward misspecification errors cannot aid in the prediction of fund returns and as a result there then appears to be little relation between forecasted and realized alphas.

This paper has documented that indeed the standard procedure of using a five year rolling window to estimate fund alphas and betas tends to place funds with questionable parameter estimates in the extreme deciles. Sorting funds by their predicted betas induces an inverse sort on alphas. This indicates the poor beta estimates are feeding back into poor alpha estimates. When the OLS model underestimates beta it tries to make up for this by then overestimating alpha. Of course, none of this helps predict future returns or even future factor loadings. The latter can be seen by looking at the stability of the out of sample beta estimates. As shown here very high and low alpha sorted deciles produce the least accurate beta forecasts going forward.

This paper suggests that it may be useful to depart from the research methodology in which a single model is used across all funds. Instead one might wish to see if a model has done a good job of predicting a particular fund's excess returns in the past prior to using that model on that fund in the future. This naturally leads to another possibility; using a variety of different models might allow for the identification of yet more funds whose performance can be predicted out of sample. Indeed the results presented here indicate that an alternative model specification that allows for time variation in a fund manager's alpha and beta can be gainfully employed towards exactly this end.

The simplest system for back checking a model is to require that it accurately predict the sign of a fund's abnormal return in month $t-1$ prior to allowing it to make predictions about the fund in month t. Remarkably even this very crude back test is enough to produce alpha sorted portfolios for which the top *two* deciles yield positive and statistically significant above market returns. Depending on the model used one can obtain risk adjusted (using a 4-factor model for the out-of-sample risk adjustment) above market returns of about 3.5% to 7.0% per year. These results also hold up under an eight factor model designed to control for stale pricing issues. The resulting portfolios picked by the different models are also very diverse with a common fund overlap of only about a third. This means that for every 100 funds selected by either model for inclusion in its top decile only about 33 of these funds will appear on both lists.

The results in this paper indicate that if one is looking to see if managerial talent can produce above market returns then it may pay to experiment with a number of statistical models. Mutual funds can, at times, follow dynamically complex strategies. It is unreasonable to expect a single model to accurately track the resulting changes in factor loadings without then producing large misspecification errors. By using a variety of models and requiring them to prove their accuracy before employing them on a particular fund, it may be possible to identify far more funds with either positive or negative out-of-sample performance than has previously been thought.

References

Admati, A., 1985, "A Noisy Rational Expectations Equilibrium for Multi-asset Securities Markets," *Econometrica*, 53, 629-657.

Avramov, D, and R. Wermers, 2005, "Investing in Mutual Funds when Returns are Predictable," forthcoming *Journal of Financial Economics*.

Baks, K., A. Metrick, and J. Wachter, 2001, "Should Investors Avoid All Actively Managed Funds? A Study in Bayesian Performance Evaluation," *Journal of Finance*, 56(1), 45-86.

Berk, J., 1995, "A Critique of Size-Related Anomalies," *Review of Financial Studies*, 8(2), 275-286.

Berk, J., and R. Green, 2004, "Mutual Fund Flows and Performance in Rational Markets." forthcoming *Journal of Political Economy*.

Berk, J., and J. Xu, 2004, "Persistence and Fund Flows of the Worst Performing Mutual Funds," U.C. Berkeley working paper, http://faculty.haas.berkeley.edu/berk/persist.html.

Blake, D., B. Lehmann, and A. Timmermann, 1999, "Asset Allocation Dynamics and Pension Fund Performance," *Journal of Business*, 72(4), 429-461.

Bollen, N., and J. Busse, 2004, "Short-Term Persistence in Mutual Fund Performance," *Review of Financial Studies*, 18(2), 569-597.

Bossaerts, P., and P. Hillion, 1999, "Implementing Statistical Criteria to Select Return Forecasting Models: What do we Learn?," *Review of Financial Studies*, 12(2), 405-428.

Brown, S., and W. Goetzmann, 1995,"Performance Persistence," *Journal of Finance*, 50(2), 679-698.

Brown, S., and W. Goetzmann, 1997, "Mutual Fund Styles," *Journal of Financial Economics*, 43(3), 373-399.

Busse, J. and P. Irvine, 2005, "Bayesian Alphas and Mutual Fund Persistence," forthcoming *Journal of Finance*.

Carhart, M., 1995, "Survivor Bias and persistence in Mutual Fund Performance," Ph.D dissertation, University of Chicago Graduate School of Business, http://wwwlib.umi.com/dissertations/dlnow/9609921.

Carhart, M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52(1), 57-82.

Chalmers, J., R. Edelen, and G. Kadlec, 2001, "On the Perils of Financial Intermediaries Setting Security Prices: The Mutual Fund Wild Card Option," *Journal of Finance*, 56(6), 2209-36.

Cohen, R., J. Coval, and Ľ. Pástor, 2005, "Judging Fund Managers by the Company they Keep," *Journal of Finance*, 60(3), 1057-1096.

Elton, E., M. Gruber, and C. Blake, 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(6), 2415-2430.

Ferson, W., and C. Harvey, 1991, "The Variation of Economic Risk Premiums," *Journal of Political Economy*, 99, 385-415.

Ferson, W., and C. Harvey, 1993, "The Risk and Predictability of International Equity Returns," *Review of Financial Studies*, 6, 527-566.

Ferson, W., and K. Khang, 2002, "Conditional Performance Measurement Using Portfolio Weights: Evidence for Pension Funds." *Journal of Financial Economics*, 65, 249-282.

Ferson, W., and R. Korajczyk, 1995, "Do Arbitrage Pricing Models Explain the Predictability of Stock Returns?," *Journal of Business*, 68(3), 309-49.

Ferson, W., and R. Schadt, 1996, "Measuring Fund Strategy and Performance in Changing Economic Conditions," *Journal of Finance*, 51(2), 425-462.

Goetzmann, W., Z. Ivkovic, and G. Rouwenhorst, 2001, "Day Trading International Mutual Funds: Evidence and Policy Solutions," *Journal of Financial and Quantitative Analysis*, 36(3), 287-310.

Goyal, A., and I. Welch, 2006, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," forthcoming *Review of Financial Studies.*

Greene, J., and C. Hodges, 2002, "The Dilution Impact of Daily Fund Flows on Open-End Mutual Funds," *Journal of Financial Economics*, 65(1), 131-158.

Grinblatt, M., and S. Titman, 1989, "Portfolio Performance Evaluation: Old Issues and New Insights," *Review of Financial Studies*, 2(3), 396-422.

Grinblatt, M., and S. Titman, 1994, "A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques," *Journal of Financial and Quantitative Analysis*, 29(3), 419-444.

Hendricks, D., J. Patel, and R. Zeckhauser, 1993, "Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974-1988," *Journal of Finance*, 48(1), 93-130.

Harvey, A., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, U.K..

Kacperczyk, M., C. Sialm, and L. Zheng, 2005a, “On the Industry Concentration of Actively Managed Equity Mutual Funds,” *Journal of Finance*, 60(4), 1983-2011.

Kacperczyk, M., C. Sialm, and L. Zheng, 2005b, “Unobserved Actions of Mutual Funds,” University of Michigan working paper, http://webuser.bus.umich.edu/luzheng/web_page/lu_papers/unobserved.pdf.

Mamaysky, H., M. Spiegel, and H. Zhang, 2005, “Estimating the Dynamics of Mutual Fund Alphas and Betas,” Yale School of Management working paper.

Pástor, Ľ., and R. Stambaugh, 2002, “Investing in Equity Mutual Funds,” *Journal of Financial Economics*, 63(3), 351-380.

Pesaran, H., and A. Timmermann, 2005, “Real-Time Econometrics,” *Econometric Theory*, 21(1), 212-231.

Timmermann, A., and C. Granger, 2004, “Efficient Market Hypothesis and Forecasting,” *International Journal of Forecasting*, 20(1), 15-27.

White, H., 2000, “A Reality Check for Data Snooping,” *Econometrica*, 68(5), 1097-1126.

Zitzewitz, E., 2003, “Who Cares about Shareholders? Arbitrage-Proofing Mutual Funds,” *Journal of Law, Economics, and Organization*, 19(2), 245-280.

7 Appendix: A Dynamic Kalman Filter Model of Mutual Fund Portfolios

A detailed derivation of the Kalman filter model presented here can be found in Mamaysky, Spiegel, and Zhang (2002). Thus, to conserve space this Appendix presents just the necessary details for carrying out the estimation procedure.

Returning to equation (1), for uninformed individuals the α_t terms always equal zero. However, for investors possessing information the α_t's may at times be positive or negative. This view of asset returns is in line with heterogeneous information models such as Admati (1985). Thus, to be precise the α_t terms should have a subscript indicating the individual and his information. For notational simplicity these indicators are suppressed, but one should keep in mind that the return equations are conditional on an investor's information set.

Even if the factor loadings for stocks and bonds do not change over time it is unlikely that this will be true of any actively managed portfolio containing these same instruments.[16] Let w_{it} represent the fraction of the portfolio in security i at time t, and W_t the $I \times 1$ vector containing the I individual weights. Then the portfolio's time t return equals the weighted average of the returns from the underlying I assets:

$$r_{Pt} - r_{ft} = W_{t-1}{}'\Big(\alpha_t + \beta'(r_{mt} - r_{ft}) + \epsilon_t\Big) - k_t. \tag{3}$$

The β term represents a matrix with I columns containing the vectors β_i. The k_t term equals the transactions costs incurred by the portfolio, which for mathematical tractability are assumed to be proportional to the funds under management.

Absent information about a fund's holdings as well as the alphas and betas of the underlying assets the parameters in equation (3) cannot be estimated. However, these problems can be overcome by adding some additional assumptions that obviate the need to know the underlying portfolio's composition.

Let F_t represent some signal (normalized to have an unconditional mean of zero) that a particular fund uses to trade. Now assume that the signal's value follows the AR(1) process

$$F_t = \gamma_F F_{t-1} + \eta_t \tag{4}$$

through time. The $\gamma_F \in [0, 1)$ coefficient measures the degree to which the signal's value persists over time, and η_t represents an i.i.d. innovation.

If the signal F has value then one expects it to influence both the fund's holdings, and future expected stock returns. Statistically, these dual impacts can be represented by assuming that the portfolio weights follow:

$$w_{it} = \bar{w}_i + l_i F_t, \tag{5}$$

[16] Many studies like those of Ferson and Harvey (1991, and 1993), and Ferson and Korajczyk (1995) question whether or not individual security loadings are constant (the β_i term in (1)). However, this will not qualitatively alter this paper's conclusion that fund loadings change over time. If anything such underlying intertemporal variation in the underlying securities will only add to the importance of allowing for time variation in the mutual funds themselves.

and that stock alphas equal

$$\alpha_{it} = \bar{\alpha}_i F_{t-1}. \tag{6}$$

Here $\bar{w}_i$ represents the steady-state fraction of the strategy invested in a given security. Alternatively, $\bar{w}_i$ can depend upon any set of observable variables, in which case it may be time dependent. The variable l_i is stock i's loading on a common unobservable factor F_t which shifts the portfolio weights from their steady-state values. One can view this formulation as an empirical application of Admati's (1985) general equilibrium asset pricing model with asymmetric information. It is also generally consistent with Blake, Lehmann, and Timmermann's (1999) finding of mean reversion in fund weightings across securities among UK pension funds. Finally, $\bar{\alpha}_i$ represents the degree to which a stock's expected return is predictable by the signal F. If the signal has no value then all of the $\bar{\alpha}_i$ terms equal zero. Also, the present specification insures that the steady state alpha values equal zero.

Now use (5), and (6) in (3) to produce the Kalman observation equation:

$$\begin{aligned} r_{Pt} - r_{ft} &= (\bar{W} + lF_{t-1})'(\bar{\alpha}F_{t-1} + \beta'(r_{mt} - r_{ft} + \epsilon_t) - k_t \qquad (7) \\ &= l'\bar{\alpha}F_{t-1}^2 - k_t + \bar{W}'\beta'(r_{mt} - r_{ft}) + (\bar{W}'\bar{\alpha} + l'\beta'(r_{mt} - rft))F_{t-1} + (\bar{W} + lF_{t-1})'\epsilon_t \\ &= b_P F_{t-1}^2 - k_t + \bar{\beta}_P(r_{mt} - rft) + (\bar{a}_P + c_P(r_{mt} - rft))F_{t-1} + \epsilon_{Pt}. \end{aligned}$$

In this equation b_P equals $l'\bar{\alpha}$, $\bar{\beta}_P$ equals $\bar{W}'$, $\bar{\alpha}_P$ equals $\bar{W}'\bar{\alpha}$, c_P equals $l'\beta'$, and ϵ_{Pt} equals $(\bar{W} + lF_{t-1})'\epsilon_t$.

The $\bar{\alpha}_i$, $\bar{\alpha}_P$, and b_P each play a unique economic role in the analysis. In equation (6), $\bar{\alpha}_i \neq 0$ implies that a given fund's signal has a systematic relationship with the instantaneous excess returns of individual stocks in the economy. Therefore, one can alternatively write $\bar{\alpha}_{iP}$ to indicate that this coefficient is both stock *and* fund dependent. The point, though, of having non-zero $\bar{\alpha}_i$'s is to allow the fund's α_P to systematically depend on the fund's trading strategy F. This dependence comes about through a linear term, the $\bar{\alpha}_P$ and a quadratic term b_P. There is no constant alpha term in α_P because in the long-run all alphas are assumed to be zero (their unconditional value). The linear term $\bar{\alpha}_P$ simply measures the degree to which a given fund's strategy is actually related to the instantaneous alphas of individual stocks. Since F can be positive or negative, a non-zero α_P does not indicate either under or over performance. The quadratic term b_P, on the other hand, does indicate exactly this – it measures the degree to which a fund is able to systematically go long (short) positive (negative) alpha stocks.[17] Note that this is a sufficient, though not necessary, condition for a given fund to exhibit occasional (as opposed to systematic) risk-adjusted outperformance. A weaker and necessary condition is that a fund's α_P is persistent and occasionally positive (which obtains when $\bar{\alpha}_P \neq 0$ and when $\gamma_F > 0$).

The empirical model derived above is very flexible. For example, if one assumes that η_t has a variance of zero, or that γ_F equals zero the Ferson and Schadt (1996) specification can

[17]Intuitively, b_P can be thought of as the covariance between a fund's security weights ($f(t)$) and the underlying security alphas.

be reproduced. Importantly, however, even absent these assumptions the model can still be estimated. Also note that nowhere does the econometrician need data on the actual portfolio weights used to produce the observed returns.

Due to the F^2_{t-1} term in (7) the standard Kalman filtering techniques will fail as the conditional variance of $r_P(t) - r(t)$ will no longer be independent of the estimated values of F_{t-1}. The standard solution is to use a first-order Taylor expansion around the conditional expectation of F_{t-1}, or

$$\begin{aligned} F^2_{t-1} &\approx 2\,\mathbb{E}\Big[F_{t-1}\Big|r_{P,t-1} - r_{t-1}, F_{t-2}\Big]F_{t-1} \\ &\quad -\mathbb{E}\Big[F_{t-1}\Big|r_{P,t-1} - r_{t-1}, F_{t-2}\Big]^2 \end{aligned} \tag{8}$$

to replace the F^2_{t-1} term in equation (7) where $\mathbb{E}$ is the expectations operator.[18] Equation (4) then forms the state equation.[19] Note, the vector c_P has n elements (one for each risk factor) but only n-1 degrees of freedom. Thus, in the scalar case (as in the CAPM) it can be normalized to one when estimating the model. In the case where n is greater than one, at least one element's value must be fixed or some other normalization must be applied. The other fact needed for estimation is that the variance of $\epsilon_p(t)$, conditional on time $t-1$ information, is given by

$$\text{Var}_{t-1}\Big(\epsilon_{Pt}\Big) = \sum_{i=1}^{I} w^2_{i,t-1}\text{Var}_{t-1}\Big(\epsilon_{it}\Big).$$

This follows from $\epsilon_{Pt} = W'_{t-1}\epsilon_t$, and from the fact that all ϵ_{it}'s are independent.

The system specified in equations (7) and (4) imbeds an important timing convention. The alphas and betas which determine time t returns are known at time $t-1$ (assuming that k_t is deterministic). Therefore any covariance which exists between a portfolio's time t alphas and time t market returns indicates an ability of the portfolio manager to make investment decisions at time $t-1$ which successfully anticipate market returns at time t. Similarly for time t betas and time t market returns.

[18] For details about extended Kalman filtering see Harvey (1989).

[19] The estimated dynamic Kalman filter model bears some philosophical resemblance to the Bayesian approaches found in Baks, Metrick, and Wachter (2001), and Pástor and Stambaugh (2002).

Table 1: **Distribution of Estimated Alphas and Betas.** This table sorts all domestic equity funds into 10 deciles based on estimated 1-factor or 4-factor OLS alphas and market betas. In month t, alphas and betas are estimated from a rolling regression based on t-60 to t-1 returns. All monthly alpha and beta estimates from January 1970 to December 2002 are pooled prior to sorting. The first two columns report the mean estimated beta value for CAPM-beta sorted deciles, and the corresponding mean alpha value within each decile. The next two columns repeat the excise for the Carhart 4-factor estimated alpha and a return-weighted beta. The return-weighted beta is defined as $\beta_i{}' = (\beta_{i1}\bar{r}_1 + \beta_{i2}\bar{r}_2 + \beta_{i3}\bar{r}_3 + \beta_{i4}\bar{r}_4)/(\bar{\beta}_1\bar{r}_1 + \bar{\beta}_2\bar{r}_2 + \bar{\beta}_3\bar{r}_3 + \bar{\beta}_4\bar{r}_4)$, where r_j and β_{ij} refer to the j^{th} factor and the corresponding factor loading for fund i, and $\bar{r}_j$ and $\bar{\beta}_j$ refer to their time-series and cross-sectional means, respectively. The last six columns flip the sorting sequence: deciles are sorted on alpha values. The mean alpha value for each decile is reported in the "Sorted α" column, and the corresponding mean beta value is reported in the mean β columns. The Std α columns report the average standard error of alphas within each alpha-sorted portfolio.

Decile	CAPM Sorted β	CAPM mean α	Carhart Sorted β'	Carhart mean α	CAPM Sorted α	CAPM mean β	CAPM Std α	Carhart Sorted α	Carhart mean β'	Carhart Std α
1	0.4000	0.0007	-1.9819	0.0014	-0.0086	1.0788	0.0050	-0.0074	0.6963	0.0044
2	0.6263	0.0002	0.5542	0.0002	-0.0037	0.9786	0.0033	-0.0036	0.9229	0.0027
3	0.7359	0.0000	0.6881	0.0000	-0.0022	0.9254	0.0028	-0.0024	0.9250	0.0023
4	0.8275	-0.0002	0.7965	-0.0002	-0.0013	0.9034	0.0026	-0.0016	0.8735	0.0022
5	0.9010	-0.0004	0.8997	-0.0004	-0.0005	0.8914	0.0024	-0.0008	0.8718	0.0021
6	0.9596	-0.0003	1.0073	-0.0006	0.0002	0.8726	0.0024	-0.0002	0.8597	0.0020
7	1.0177	-0.0007	1.1298	-0.0008	0.0009	0.8680	0.0026	0.0005	0.8006	0.0020
8	1.0992	-0.0010	1.2843	-0.0010	0.0018	0.8774	0.0029	0.0013	0.6956	0.0022
9	1.2197	-0.0012	1.4925	-0.0014	0.0031	0.8843	0.0035	0.0026	0.7164	0.0027
10	1.4856	-0.0005	1.9588	-0.0025	0.0069	0.9932	0.0049	0.0064	0.4766	0.0042

Table 2: **Beta Errors in Sorted Alpha Deciles.** This table first calculates monthly OLS alphas and betas in month t from a rolling regression based on t-60 to t-1 returns. These estimated parameters are then used as forecasts in month t. This process is used over a twelve month period and the average forecasted alpha and beta is calculated. Next the realized fund returns for the twelve month period is regressed on the market model to produce an in sample alphas and beta for the period in question. Let $\Delta\beta_{MKT}$ equal the difference between the predicted and realized beta ($\Delta\beta_{MKT} \equiv \beta_{MKT,predicted} - \beta_{MKT,realized}$) . Similarly, for the four factor model the difference between the predicted and realized factor loadings will be referred to as $\Delta\beta_i$ for i equal to MKT, SMB, HML, and MOM. All of the mean alpha values and beta differences are then pooled across the different months for the testing period of January 1970 to December 2002. Ten deciles are then created based upon the estimated alphas. The table reports by decile the mean alpha value, the parameters from the following regression: $\Delta\beta_i = c_i + error$, and the standard deviation of $\Delta\beta_i$ ($i = MKT, SMB, HML, andMOM$) within each decile. $T_{NW,11}(c_{MKT})$ are the Newey-West T-statistics for c_{MKT} with 11 lags. $T_{OLS}(c_{MKT})/\sqrt{12}$ column reports the T-statistics as the OLS T-statistics divided by the square-root of 12.

Deciles	Sorted α	c_{MKT}	$T_{NW,11}(c_{MKT})$	$T_{OLS}(c_{MKT})/\sqrt{12}$	Std($\Delta\beta_{MKT}$)	Std($\Delta\beta_{SMB}$)	Std($\Delta\beta_{HML}$)	Std($\Delta\beta_{MOM}$)
	A. CAPM							
1	-0.0077	0.0508	(16.37)	(6.02)	0.34			
2	-0.0033	0.0188	(10.78)	(3.15)	0.24			
3	-0.0021	0.0160	(9.81)	(2.87)	0.22			
4	-0.0012	0.0095	(6.24)	(1.77)	0.22			
5	-0.0005	0.0052	(3.77)	(1.08)	0.20			
6	0.0001	0.0016	(1.15)	(0.33)	0.19			
7	0.0008	-0.0050	(-3.53)	(-1.03)	0.20			
8	0.0015	-0.0119	(-7.85)	(-2.28)	0.21			
9	0.0027	-0.0247	(-14.97)	(-4.34)	0.23			
10	0.0059	-0.0388	(-19.47)	(-5.80)	0.27			
	B. Carhart 4-factor model							
1	-0.0067	-0.0631	(-16.93)	(-5.60)	0.48	0.58	0.68	0.58
2	-0.0033	-0.0274	(-13.63)	(-3.92)	0.30	0.34	0.48	0.36
3	-0.0022	-0.0262	(-14.17)	(-4.12)	0.27	0.31	0.43	0.34
4	-0.0014	-0.0195	(-10.90)	(-3.21)	0.26	0.29	0.40	0.31
5	-0.0007	-0.0226	(-13.94)	(-4.01)	0.24	0.27	0.37	0.29
6	-0.0001	-0.0224	(-13.84)	(-4.06)	0.23	0.27	0.37	0.29
7	0.0005	-0.0162	(-9.92)	(-2.87)	0.24	0.28	0.37	0.29
8	0.0013	-0.0263	(-15.26)	(-4.43)	0.25	0.31	0.40	0.32
9	0.0025	-0.0272	(-12.99)	(-3.76)	0.31	0.40	0.46	0.39
10	0.0060	-0.0631	(-20.59)	(-6.47)	0.41	0.49	0.62	0.55

Table 3: **Out of Sample Performance (No Parameter Forecast Filters, No Back Testing.** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting from January 1970, the 1-factor and 4-factor OLS models (OLS 1 and OLS 4) and 1-factor and 4-factor Kalman models (Kal 1 and Kal 4) independently sort funds into equally weighted portfolios based on forecasted alphas. Decile 10 contains funds with the highest forecasted alphas. The "Top X" portfolios contain an equally weighted portfolio of the X funds with the highest alphas. Each portfolio is held for 1 month and then rebalanced until December 2002. There are no restrictions on forecasted alphas and betas. Panel A reports the 4-factor (MKT, SMB, HML, and MOM) adjusted monthly returns realized during the entire period and the corresponding T-statistics. In Panel B, the portfolio returns are risk-adjusted by both the four factors and one-month-lagged 4-factor returns (8 factors in total).

Decile	OLS 1	OLS 4	Kal 1	Kal 4	OLS 1	OLS 4	Kal 1	Kal 4
	A1. Four Factor Monthly Alpha				A2. T-ratio			
1	-0.0009	-0.0018	-0.0015	-0.0019	(-1.15)	(-2.68)	(-1.97)	(-2.91)
2	-0.0000	-0.0011	-0.0009	-0.0009	(-0.08)	(-2.68)	(-1.73)	(-1.93)
3	-0.0002	-0.0006	-0.0005	-0.0008	(-0.37)	(-1.75)	(-1.13)	(-2.03)
4	-0.0004	-0.0007	-0.0001	-0.0007	(-0.97)	(-1.80)	(-0.14)	(-1.88)
5	-0.0003	-0.0005	-0.0003	-0.0003	(-0.70)	(-1.33)	(-0.89)	(-0.99)
6	-0.0006	-0.0003	-0.0002	-0.0004	(-1.79)	(-0.95)	(-0.63)	(-1.32)
7	-0.0004	-0.0005	-0.0005	-0.0002	(-1.03)	(-1.56)	(-1.39)	(-0.64)
8	-0.0001	-0.0004	-0.0008	0.0001	(-0.19)	(-0.89)	(-1.78)	(0.33)
9	-0.0003	0.0004	0.0000	0.0005	(-0.50)	(0.83)	(0.07)	(1.21)
10	-0.0008	0.0016	0.0005	0.0005	(-1.14)	(2.39)	(0.67)	(0.73)
Top 20	-0.0009	0.0030	0.0014	0.0001	(-0.96)	(2.64)	(1.11)	(0.15)
Top 10	-0.0005	0.0031	0.0013	-0.0004	(-0.42)	(2.18)	(0.80)	(-0.33)
Top 5	-0.0015	0.0037	0.0022	-0.0002	(-0.76)	(2.04)	(1.09)	(-0.14)
	B1. Eight Factor Monthly Alpha				B2. T-ratio			
1	-0.0009	-0.0017	-0.0010	-0.0016	(-1.13)	(-2.45)	(-1.27)	(-2.34)
2	-0.0000	-0.0009	-0.0010	-0.0008	(-0.08)	(-2.06)	(-1.84)	(-1.63)
3	-0.0000	-0.0004	-0.0004	-0.0006	(-0.08)	(-0.95)	(-0.90)	(-1.51)
4	-0.0003	-0.0006	-0.0000	-0.0006	(-0.59)	(-1.42)	(-0.08)	(-1.59)
5	-0.0004	-0.0003	-0.0003	-0.0003	(-0.87)	(-0.90)	(-0.89)	(-0.89)
6	-0.0005	-0.0002	-0.0002	-0.0004	(-1.35)	(-0.64)	(-0.61)	(-1.07)
7	-0.0002	-0.0005	-0.0005	-0.0002	(-0.48)	(-1.48)	(-1.28)	(-0.64)
8	0.0001	-0.0004	-0.0007	0.0002	(0.11)	(-1.04)	(-1.64)	(0.44)
9	-0.0001	0.0005	0.0002	0.0005	(-0.25)	(0.96)	(0.37)	(1.10)
10	-0.0007	0.0015	0.0006	0.0005	(-0.89)	(2.16)	(0.76)	(0.76)
Top 20	-0.0007	0.0026	0.0011	0.0000	(-0.70)	(2.19)	(0.83)	(0.00)
Top 10	-0.0007	0.0024	0.0011	-0.0005	(-0.47)	(1.57)	(0.63)	(-0.34)
Top 5	-0.0022	0.0029	0.0023	-0.0001	(-1.05)	(1.53)	(1.06)	(-0.08)

Table 4: **Out of Sample Performance With Various α and β Filters (No Back Testing).** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting from January 1970 until December 2002, the 1-factor and 4-factor OLS models (OLS 1 and OLS 4), and the 1-factor and 4-factor Kalman models (Kal 1 and Kal 4) independently sort funds into 10 deciles based on their forecasted alphas. Decile 10 contains funds with the highest forecasted alphas. For each model, 10 equally-weighted portfolios are constructed from stocks within the 10 deciles. Each portfolio is held for 1 month and rebalanced at the beginning of the next month. For any fund to be included in any decile the predicted alpha and beta values must lie within the specified boundary. In Panels A and C, the alpha boundaries are fixed at ±2% per month and the beta range changes. In Panels B and D, the alpha boundary changes but the beta range is fixed at 0 to 2. The table reports the 4-factor (MKT, SMB, HML, and MOM) adjusted monthly returns realized in the entire period for Decile 10 funds (those expected to generate the highest future return).

		OLS 1	OLS 4	Kal 1	Kal 4	OLS 1	OLS 4	Kal 1	Kal 4
A. Decile 10 Performance with various beta ranges.									
β_{Min}	β_{Max}	Four Factor Monthly Alpha				T-ratio			
0.80	1.20	0.0004	0.0020	0.0005	0.0010	(0.70)	(2.79)	(0.93)	(1.84)
0.60	1.40	-0.0000	0.0019	0.0009	0.0008	(-0.03)	(2.86)	(1.49)	(1.42)
0.40	1.60	-0.0003	0.0017	0.0008	0.0007	(-0.42)	(2.69)	(1.27)	(1.33)
0.20	1.80	-0.0007	0.0016	0.0007	0.0007	(-1.03)	(2.34)	(1.05)	(1.25)
0.00	2.00	-0.0008	0.0016	0.0007	0.0007	(-1.22)	(2.35)	(1.04)	(1.28)
B. Decile 10 Performance with various α ranges.									
α_{Min}	α_{Max}	Four Factor Monthly Alpha				T-ratio			
-0.02	0.02	-0.0008	0.0016	0.0007	0.0007	(-1.22)	(2.35)	(1.04)	(1.28)
-0.04	0.04	-0.0010	0.0016	0.0005	0.0005	(-1.39)	(2.37)	(0.74)	(0.85)
-0.06	0.06	-0.0010	0.0016	0.0007	0.0006	(-1.39)	(2.37)	(0.91)	(0.95)
-0.08	0.08	-0.0010	0.0016	0.0006	0.0005	(-1.39)	(2.37)	(0.87)	(0.80)
-0.10	0.10	-0.0010	0.0016	0.0006	0.0004	(-1.39)	(2.37)	(0.77)	(0.74)
-1.00	1.00	-0.0010	0.0016	0.0004	0.0005	(-1.42)	(2.39)	(0.55)	(0.75)

Table 5: **Out of Sample Performance (With Parameter Forecast Filters, With Back Testing).** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting from January 1970 until December 2002, the 1-factor and 4-factor OLS models (OLS 1 and OLS 4), and 1-factor and 4-factor Kalman models (Kal 1 and Kal 4) independently sort funds into 10 deciles based on forecasted alphas. Decile 10 contains funds with the highest forecasted alphas. For each model, 10 equal-weighted portfolio are constructed from stocks within the 10 deciles. Each portfolio is held for 1 month and rebalanced at the beginning of the next month. For any fund to be included in any decile the following must be true: 1) the absolute value of forecasted alpha must be less than 2% per month, 2) the beta must be greater than 0 but less than 2, and 3) in the previous month the out of sample forecasted alpha and the difference between the realized excess return and the market return must have the same sign. Finally, the table also constructs model by model equally-weighted portfolios containing the 20, 10 and 5 funds with the highest alphas forecasts. Panels A reports the monthly excess return and Sharpe ratio for each portfolio after regressing the sequence of returns on the appropriate factor model. Panels B reports the 4-factor (MKT, SMB, HML, and MOM) adjusted monthly returns realized during the entire period and the corresponding T-statistics. Though not reported, the rank correlation between decile number and the average realized excess return for each model is highly significant (correlation > 0.9, P-value < 0.01).

Decile	OLS 1	OLS 4	Kal 1	Kal 4	OLS 1	OLS 4	Kal 1	Kal 4
	A1. Monthly Return Above the Risk Free Rate				A2. Monthly Sharpe Ratio			
1	0.0011	0.0008	0.0013	0.0016	0.0215	0.0165	0.0266	0.0318
2	0.0020	0.0020	0.0026	0.0027	0.0402	0.0410	0.0539	0.0584
3	0.0021	0.0026	0.0027	0.0027	0.0444	0.0557	0.0565	0.0576
4	0.0032	0.0028	0.0036	0.0026	0.0702	0.0607	0.0803	0.0582
5	0.0042	0.0034	0.0033	0.0033	0.0987	0.0774	0.0757	0.0775
6	0.0045	0.0037	0.0043	0.0038	0.1062	0.0864	0.0996	0.0896
7	0.0052	0.0043	0.0052	0.0050	0.1245	0.1003	0.1225	0.1159
8	0.0052	0.0056	0.0050	0.0055	0.1228	0.1254	0.1179	0.1243
9	0.0059	0.0061	0.0064	0.0058	0.1318	0.1322	0.1427	0.1285
10	0.0070	0.0076	0.0078	0.0066	0.1474	0.1508	0.1706	0.1378
Top 20	0.0066	0.0074	0.0078	0.0057	0.1337	0.1343	0.1636	0.1159
Top 10	0.0078	0.0084	0.0084	0.0062	0.1494	0.1421	0.1769	0.1232
Top 5	0.0080	0.0078	0.0089	0.0063	0.1482	0.1245	0.1804	0.1186
	B1. Four Factor Monthly Alpha				B2. T-ratio			
1	-0.0026	-0.0037	-0.0023	-0.0031	(-2.88)	(-3.90)	(-2.57)	(-3.83)
2	-0.0020	-0.0025	-0.0016	-0.0016	(-2.72)	(-3.79)	(-2.51)	(-2.55)
3	-0.0021	-0.0018	-0.0015	-0.0018	(-3.38)	(-3.24)	(-2.45)	(-3.25)
4	-0.0010	-0.0016	-0.0005	-0.0019	(-1.73)	(-2.75)	(-1.00)	(-3.93)
5	0.0000	-0.0009	-0.0010	-0.0009	(0.04)	(-1.55)	(-2.04)	(-1.86)
6	-0.0000	-0.0009	-0.0003	-0.0005	(-0.00)	(-1.77)	(-0.57)	(-1.08)
7	0.0007	-0.0001	0.0007	0.0004	(1.09)	(-0.15)	(1.46)	(0.93)
8	0.0004	0.0008	0.0005	0.0010	(0.64)	(1.37)	(0.83)	(1.88)
9	0.0009	0.0016	0.0013	0.0013	(1.30)	(2.53)	(2.03)	(2.43)
10	0.0021	0.0037	0.0031	0.0023	(2.35)	(4.03)	(3.95)	(3.23)
Top 20	0.0019	0.0046	0.0031	0.0020	(2.01)	(3.93)	(3.57)	(2.45)
Top 10	0.0030	0.0057	0.0036	0.0027	(2.48)	(3.91)	(3.70)	(2.59)
Top 5	0.0032	0.0048	0.0040	0.0022	(2.20)	(2.73)	(3.34)	(1.81)

Table 6: **Eight Factor Model Out of Sample Performance (With Parameter Forecast Filters, With Back Testing).** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting from January 1970 until December 2002, the 1-factor and 4-factor OLS models (OLS 1 and OLS 4), and 1-factor and 4-factor Kalman models (Kal 1 and Kal 4) independently sort funds into 10 deciles based on forecasted alphas. Decile 10 contains funds with the highest forecasted alphas. For each model 10 equally-weighted portfolio are constructed from stocks within the 10 deciles. At the beginning of month t, funds are sorted according to predicted alphas in month t. Each portfolio is held for 1 month and rebalanced at the beginning of the next month. For any fund to be included in any decile 1) the absolute value of forecasted alpha must be less than 2% a month, 2) the beta must be greater than 0 but less than 2, and 3) the forecasted alpha and the difference between the realized excess return and the market return in the previous month must have the same sign. Finally, the table also constructs equal-weighted portfolios containing the 20, 10 and 5 funds with the highest alphas forecast by each model. The eight factor model contains the four Carhart factors plus their one month lagged values.

Decile	OLS 1	OLS 4	Kal 1	Kal 4	OLS 1	OLS 4	Kal 1	Kal 4
	Eight Factor Monthly Alpha				T-ratio for lagged-factor adjusted return			
1	-0.0024	-0.0030	-0.0017	-0.0026	(-2.56)	(-3.04)	(-1.80)	(-3.06)
2	-0.0016	-0.0019	-0.0012	-0.0013	(-2.08)	(-2.80)	(-1.78)	(-1.97)
3	-0.0017	-0.0014	-0.0013	-0.0013	(-2.63)	(-2.31)	(-2.04)	(-2.26)
4	-0.0009	-0.0011	-0.0004	-0.0017	(-1.52)	(-1.82)	(-0.68)	(-3.37)
5	-0.0000	-0.0005	-0.0009	-0.0008	(-0.07)	(-0.84)	(-1.77)	(-1.61)
6	-0.0001	-0.0009	-0.0004	-0.0007	(-0.20)	(-1.63)	(-0.91)	(-1.38)
7	0.0005	-0.0002	0.0007	0.0003	(0.72)	(-0.36)	(1.30)	(0.65)
8	0.0002	0.0007	0.0002	0.0009	(0.34)	(1.15)	(0.29)	(1.51)
9	0.0004	0.0014	0.0011	0.0011	(0.62)	(2.15)	(1.59)	(1.93)
10	0.0016	0.0032	0.0027	0.0018	(1.75)	(3.29)	(3.17)	(2.46)
Top 20	0.0014	0.0038	0.0029	0.0016	(1.39)	(3.12)	(3.10)	(1.87)
Top 10	0.0025	0.0048	0.0033	0.0021	(1.92)	(3.14)	(3.15)	(1.91)
Top 5	0.0021	0.0036	0.0038	0.0016	(1.38)	(1.97)	(2.98)	(1.29)

Table 7: **Out of Sample Performance (No Parameter Forecast Filters, With Back Testing).** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting in January 1970 until December 2002, the 1-factor and 4-factor OLS models (OLS 1 and OLS 4), and 1-factor and 4-factor Kalman models (Kal 1 and Kal 4) independently sort funds into 10 deciles based on forecasted alphas. Decile 10 contains funds with highest forecasted alphas. For each model, 10 equally-weighted portfolios are constructed from the stocks within the 10 deciles. Each portfolio is then held for 1 month and rebalanced at the beginning of the next month. For any fund to be included in any decile in the previous month the out of sample forecasted alpha and the difference between the realized excess return and the market return must have the same sign. Also constructed are equally-weighted portfolios containing the 20, 10, and 5 funds with the highest alpha forecasts by each model. Panel A reports the excess returns from the four factor model, and Panel B from the eight factor model.

Decile	OLS 1	OLS 4	Kal 1	Kal 4	OLS 1	OLS 4	Kal 1	Kal 4
	A1. Four Factor Monthly Alpha				A2. T-stat			
1	-0.0017	-0.0024	-0.0012	-0.0017	(-1.20)	(-2.78)	(-1.49)	(-2.12)
2	-0.0017	-0.0014	-0.0012	-0.0013	(-1.67)	(-2.27)	(-2.05)	(-2.34)
3	-0.0010	-0.0010	-0.0006	-0.0012	(-2.90)	(-1.75)	(-1.13)	(-2.45)
4	-0.0013	-0.0010	-0.0003	-0.0013	(-1.83)	(-1.93)	(-0.59)	(-2.87)
5	-0.0003	-0.0007	-0.0012	-0.0009	(-1.06)	(-1.22)	(-2.72)	(-1.99)
6	-0.0004	-0.0011	-0.0006	-0.0006	(-0.54)	(-2.08)	(-1.33)	(-1.34)
7	-0.0000	-0.0009	0.0007	0.0002	(-0.29)	(-1.73)	(1.43)	(0.32)
8	0.0002	0.0004	-0.0003	0.0005	(-0.16)	(0.66)	(-0.57)	(0.95)
9	0.0010	0.0012	0.0009	0.0006	(1.09)	(1.84)	(1.32)	(1.01)
10	0.0021	0.0025	0.0016	0.0000	(0.85)	(2.67)	(1.89)	(0.06)
Top 20	0.0004	0.0030	0.0008	0.0002	(1.09)	(2.68)	(0.75)	(0.22)
Top 10	-0.0001	0.0044	-0.0004	-0.0002	(1.03)	(3.22)	(-0.31)	(-0.18)
Top 5	0.0003	0.0052	-0.0009	-0.0022	(0.66)	(2.90)	(-0.51)	(-1.34)
	B1. Eight Factor Monthly Alpha				B2. T-stat			
1	-0.0016	-0.0023	-0.0014	-0.0017	(-1.21)	(-2.55)	(-1.59)	(-2.01)
2	-0.0017	-0.0013	-0.0011	-0.0012	(-1.72)	(-2.03)	(-1.74)	(-2.07)
3	-0.0009	-0.0010	-0.0006	-0.0012	(-2.40)	(-1.68)	(-1.03)	(-2.22)
4	-0.0013	-0.0009	-0.0002	-0.0013	(-1.39)	(-1.61)	(-0.44)	(-2.67)
5	-0.0003	-0.0003	-0.0011	-0.0009	(-0.63)	(-0.60)	(-2.37)	(-1.87)
6	-0.0004	-0.0008	-0.0008	-0.0006	(-0.29)	(-1.43)	(-1.55)	(-1.18)
7	0.0002	-0.0010	0.0009	0.0000	(0.08)	(-1.72)	(1.56)	(0.03)
8	0.0003	0.0003	-0.0002	0.0004	(0.05)	(0.48)	(-0.37)	(0.67)
9	0.0014	0.0014	0.0009	0.0005	(1.20)	(2.09)	(1.31)	(0.74)
10	0.0026	0.0026	0.0018	0.0003	(1.19)	(2.57)	(1.96)	(0.37)
Top 20	0.0017	0.0034	0.0014	0.0003	(1.45)	(2.89)	(1.28)	(0.38)
Top 10	0.0010	0.0046	0.0003	0.0002	(1.19)	(3.19)	(0.20)	(0.13)
Top 5	0.0013	0.0053	-0.0003	-0.0024	(0.76)	(2.77)	(-0.17)	(-1.35)

Table 8: **Out of Sample Performance (Parameter Forecast Filters, With Back Testing Using Model Beta Forecasts).** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting from January 1970 until December 2002, Models 1 to 4 (the 1-factor and 4-factor OLS models, and 1-factor and 4-factor Kalman models, respectively) independently sort funds into 10 deciles based on forecasted alphas. Decile 10 contains funds with the highest forecasted alphas. For each model 10 equally-weighted portfolio are constructed from stocks within the 10 deciles. At the beginning of month t, funds are sorted according to predicted alphas in month t. Each portfolio is held for 1 month and rebalanced at the beginning of the next month. For any fund to be included in any decile 1) the absolute value of forecasted alpha must be less than 2% a month, 2) the beta must be greater than 0 but less than 2, and 3) the forecasted alpha and the risk adjusted excess return when calculated with the model's forecasted betas in the previous month must have the same sign. Finally, the table also constructs equal-weighted portfolios containing the 20, 10 and 5 funds with the highest alphas forecast by each model.

Decile	OLS 1	OLS 4	Kal 1	Kal 4	OLS 1	OLS 4	Kal 1	Kal 4
	Four Factor Monthly Alpha				T-stat			
1	-0.0021	-0.0028	-0.0023	-0.0035	(-2.41)	(-3.29)	(-2.69)	(-4.63)
2	-0.0017	-0.0022	-0.0016	-0.0011	(-2.45)	(-4.02)	(-2.52)	(-2.19)
3	-0.0015	-0.0011	-0.0013	-0.0005	(-2.42)	(-2.20)	(-2.23)	(-1.08)
4	-0.0007	-0.0012	-0.0006	-0.0006	(-1.36)	(-2.53)	(-1.28)	(-1.39)
5	-0.0003	-0.0007	-0.0008	-0.0008	(-0.46)	(-1.32)	(-1.70)	(-2.14)
6	0.0000	-0.0008	-0.0008	-0.0004	(0.08)	(-1.78)	(-1.69)	(-1.03)
7	0.0002	-0.0006	0.0003	0.0000	(0.35)	(-1.30)	(0.68)	(0.02)
8	0.0004	0.0009	-0.0000	-0.0001	(0.72)	(1.75)	(-0.09)	(-0.26)
9	0.0007	0.0009	0.0016	0.0005	(1.03)	(1.75)	(2.52)	(0.98)
10	0.0012	0.0030	0.0026	0.0014	(1.36)	(3.90)	(2.74)	(1.98)
Top 20	0.0015	0.0037	0.0024	0.0011	(1.51)	(3.34)	(2.20)	(1.33)
Top 10	0.0023	0.0041	0.0025	0.0021	(1.80)	(3.10)	(1.80)	(1.93)
Top 5	0.0024	0.0039	0.0028	0.0010	(1.42)	(2.36)	(1.53)	(0.66)

Table 9: **Out of Sample Performance Sorted by T-statistics (No Parameter Forecast Filters, No Back Testing).** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting from January 1970, the 1-factor and 4-factor OLS models (OLS 1 and OLS 4) and 1-factor and 4-factor Kalman models (Kal 1 and Kal 4) independently sort funds into equally weighted portfolios based on the in-sample t-statistics of forecasted alphas. Decile 10 contains funds with the highest t-statistics. The "Top X" portfolios contain an equally weighted portfolio of the X funds with the highest alphas. Each portfolio is held for 1 month and then rebalanced until December 2002. There are no restrictions on forecasted alphas and betas. Panel A reports the 4-factor (MKT, SMB, HML, and MOM) adjusted monthly returns realized during the entire period and the corresponding T-statistics. In Panel B, the portfolio returns are risk-adjusted by both the four factors and one-month-lagged 4-factor returns (8 factors in total).

Decile	OLS 1	OLS 4	Kal 1	Kal 4	OLS 1	OLS 4	Kal 1	Kal 4
	A1. Four Factor Monthly Alpha				A2. T-ratio			
1	0.0004	-0.0006	-0.0007	-0.0015	(0.76)	(-1.34)	(-1.28)	(-2.59)
2	-0.0001	-0.0006	-0.0003	-0.0012	(-0.11)	(-1.31)	(-0.65)	(-2.16)
3	-0.0007	-0.0007	-0.0003	-0.0007	(-1.38)	(-1.38)	(-0.55)	(-1.40)
4	-0.0005	-0.0012	-0.0000	-0.0005	(-1.00)	(-2.39)	(-0.06)	(-1.04)
5	-0.0005	-0.0008	-0.0007	-0.0005	(-1.04)	(-1.67)	(-1.79)	(-1.18)
6	-0.0004	-0.0004	-0.0008	0.0002	(-0.88)	(-0.87)	(-1.92)	(0.47)
7	-0.0003	-0.0004	-0.0002	-0.0003	(-0.61)	(-0.82)	(-0.38)	(-0.63)
8	-0.0006	0.0002	-0.0007	0.0005	(-1.25)	(0.37)	(-1.38)	(1.07)
9	-0.0005	0.0011	-0.0005	0.0004	(-1.03)	(2.25)	(-0.89)	(0.77)
10	0.0001	0.0020	-0.0003	-0.0001	(0.19)	(4.20)	(-0.46)	(-0.15)
Top 20	0.0005	0.0021	0.0001	0.0005	(0.83)	(3.67)	(0.14)	(0.87)
Top 10	0.0009	0.0027	0.0002	0.0004	(1.35)	(4.16)	(0.39)	(0.67)
Top 5	0.0010	0.0034	0.0015	0.0007	(1.27)	(4.10)	(2.15)	(0.86)
	B1. Eight Factor Monthly Alpha				B2. T-ratio			
1	0.0002	-0.0005	-0.0008	-0.0015	(0.47)	(-1.15)	(-1.32)	(-2.44)
2	0.0001	-0.0004	-0.0005	-0.0010	(0.23)	(-0.79)	(-1.04)	(-1.68)
3	-0.0007	-0.0005	-0.0001	-0.0005	(-1.38)	(-0.94)	(-0.26)	(-0.87)
4	-0.0002	-0.0009	0.0001	-0.0006	(-0.37)	(-1.79)	(0.12)	(-1.12)
5	-0.0002	-0.0007	-0.0005	-0.0004	(-0.41)	(-1.39)	(-1.38)	(-0.84)
6	-0.0002	-0.0004	-0.0008	0.0004	(-0.31)	(-0.90)	(-1.84)	(0.76)
7	-0.0001	-0.0003	0.0000	-0.0002	(-0.22)	(-0.56)	(0.09)	(-0.28)
8	-0.0006	0.0000	-0.0006	0.0007	(-1.13)	(0.01)	(-1.05)	(1.22)
9	-0.0006	0.0013	-0.0005	0.0004	(-1.12)	(2.52)	(-0.78)	(0.77)
10	0.0002	0.0019	-0.0000	-0.0001	(0.42)	(3.72)	(-0.08)	(-0.09)
Top 20	0.0004	0.0020	0.0002	0.0003	(0.76)	(3.27)	(0.32)	(0.60)
Top 10	0.0009	0.0023	0.0004	0.0002	(1.25)	(3.35)	(0.62)	(0.35)
Top 5	0.0011	0.0028	0.0019	0.0009	(1.23)	(3.31)	(2.56)	(1.08)

Table 10: **Common Fund Selections Across Models.** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting in January 1970 until December 2002, the 1-factor and 4-factor OLS models (OLS 1 and OLS 4, when subscripted O1 and O4 respectively), and 1-factor and 4-factor Kalman models (Kal 1 and Kal 4, when subscripted K1 and K4 respectively) independently sort funds into 10 deciles based on forecasted alphas. Decile 10 contains funds with highest forecasted alphas. For each model, 10 equally-weighted portfolios are constructed from the stocks within the 10 deciles. Each portfolio is then held for 1 month and rebalanced at the beginning of the next month. For any fund to be included in any decile the following must be true: 1) the absolute value of forecasted alpha must be less than 2% per month, 2) the beta must be greater than 0 but less than 2, and 3) in the previous month the out of sample forecasted alpha and the difference between the realized excess return and the market return must have the same sign. Also constructed are equally-weighted portfolios containing the 10 and 5 funds with the highest alpha forecasts by each model. Panel A reports the average ratio of firms that are selected either by Models i or j (reported as $\eta_{i,j}$). The row "9 and 10" reports the common ratio for portfolios that combine the decile 9 and 10 model-portfolios. Panel B reports the active ratio (active funds divided by all available funds) selected by each model for the selected deciles. Panel C reports the active ratio for funds selected by either Model i or Model j (reported as $D_{i,j}$).

A.	Common Ratio			
	$\eta_{O1,K1}$	$\eta_{O4,K4}$	$\eta_{O4,K1}$	$\eta_{O1,K4}$
1	0.46	0.39	0.34	0.31
2	0.25	0.20	0.17	0.16
3	0.20	0.16	0.12	0.12
4	0.18	0.13	0.10	0.09
5	0.17	0.12	0.09	0.08
6	0.17	0.11	0.09	0.07
7	0.18	0.13	0.10	0.08
8	0.20	0.17	0.13	0.12
9	0.26	0.23	0.18	0.16
10	0.43	0.38	0.34	0.29
Top 10	0.34	0.28	0.26	0.22
Top 5	0.27	0.20	0.21	0.17
9 and 10	0.35	0.31	0.26	0.23
B.	Active Ratio for Different Models			
	OLS 1	OLS 4	Kal 1	Kal 4
10	0.051	0.051	0.050	0.049
9 and 10	0.101	0.102	0.100	0.098
C.	Active Ratio Across Models			
	$D_{O1,K4}$	$D_{O4,K4}$	$D_{O4,K1}$	$D_{O1,K4}$
9	0.087	0.089	0.092	0.092
10	0.079	0.081	0.084	0.085
9 and 10	0.166	0.170	0.175	0.177

Table 11: **Out of Sample Performance: Funds Selected by Multiple Models.** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting from January 1970 until December 2002, the 1-factor and 4-factor OLS models (O1 and O4), and 1-factor and 4-factor Kalman models (K1 and K4) independently sort funds into 10 deciles based on their forecasted alphas. Decile 10 contains funds with the highest forecasted alphas. For each model, 10 equally-weighted portfolios are constructed from the stocks within the 10 deciles. For any fund to be included in any decile the following must be true: 1) the absolute value of forecasted alpha must be less than 2% per month, 2) the beta must be greater than 0 but less than 2, and 3) in the previous month the out of sample forecasted alpha and the difference between the realized excess return and the market return must have the same sign. Next, new decile portfolios $C_{i,j}$ are constructed by equally investing into funds that are commonly selected by both model i and j for each decile. Each portfolio is then held for 1 month and rebalanced at the beginning of the next month. If no common funds are available for one specific decile during month t, then this decile invests in t-bills for that month. Also constructed are equally-weighted portfolios containing the intersection of the 20, 10 and 5 funds with the highest alpha forecasts by the two models. Finally, Panel A reports monthly excess returns and the Sharpe Ratios for these portfolios. Panel B reports the 4-factor (MKT, SMB, HML, and MOM) adjusted monthly returns.

A.	Monthly Excess Return				Monthly Sharpe Ratio			
Decile	$C_{O1,K1}$	$C_{O4,K4}$	$C_{O4,K1}$	$C_{O1,K4}$	$C_{O1,K1}$	$C_{O4,K4}$	$C_{O4,K1}$	$C_{O1,K4}$
1	0.0005	-0.0002	0.0000	-0.0003	0.0104	-0.0039	0.0000	-0.0054
2	0.0025	0.0020	0.0026	0.0024	0.0479	0.0408	0.0514	0.0484
3	0.0029	0.0019	0.0026	0.0028	0.0602	0.0417	0.0557	0.0552
4	0.0039	0.0021	0.0030	0.0028	0.0851	0.0459	0.0648	0.0596
5	0.0041	0.0039	0.0040	0.0038	0.0917	0.0855	0.0915	0.0887
6	0.0044	0.0036	0.0042	0.0062	0.1061	0.0791	0.0951	0.1445
7	0.0045	0.0048	0.0043	0.0048	0.1067	0.1088	0.0996	0.1075
8	0.0037	0.0053	0.0052	0.0055	0.0855	0.1176	0.1177	0.1209
9	0.0065	0.0054	0.0056	0.0052	0.1384	0.1134	0.1189	0.1113
10	0.0078	0.0085	0.0079	0.0089	0.1669	0.1654	0.1597	0.1768
Top 20	0.0068	0.0079	0.0082	0.0077	0.1366	0.1372	0.1519	0.1428
Top 10	0.0076	0.0085	0.0079	0.0101	0.1434	0.1405	0.1399	0.1773
Top 5	0.0071	0.0079	0.0060	0.0079	0.1274	0.1325	0.1023	0.1385
B.	4-factor adjusted return				T-ratio for 4-factor adjusted return			
1	-0.0024	-0.0044	-0.0031	-0.0040	(-2.53)	(-3.78)	(-2.61)	(-3.71)
2	-0.0017	-0.0023	-0.0016	-0.0019	(-2.16)	(-2.73)	(-2.15)	(-2.38)
3	-0.0015	-0.0018	-0.0014	-0.0019	(-2.06)	(-2.75)	(-1.91)	(-2.32)
4	0.0002	-0.0018	-0.0010	-0.0013	(0.28)	(-2.72)	(-1.60)	(-1.78)
5	0.0000	-0.0002	0.0001	-0.0002	(0.07)	(-0.32)	(0.12)	(-0.28)
6	-0.0001	-0.0009	-0.0002	0.0024	(-0.22)	(-1.39)	(-0.29)	(2.74)
7	0.0003	0.0007	0.0005	0.0007	(0.45)	(0.99)	(0.71)	(0.92)
8	-0.0008	0.0007	0.0006	0.0011	(-1.17)	(1.10)	(0.85)	(1.50)
9	0.0015	0.0008	0.0012	0.0003	(1.74)	(1.11)	(1.29)	(0.31)
10	0.0038	0.0049	0.0045	0.0048	(4.02)	(5.08)	(4.38)	(4.35)
Top 20	0.0024	0.0048	0.0050	0.0036	(2.14)	(3.63)	(3.67)	(2.88)
Top 10	0.0036	0.0062	0.0041	0.0075	(2.41)	(3.92)	(2.47)	(4.53)
Top 5	0.0028	0.0054	0.0024	0.0055	(1.67)	(3.31)	(1.21)	(3.26)

Table 12: **Out of Sample Performance: Funds Selected by Only One Model.** For all domestic equity mutual funds having at least 5 years of monthly return data, both the Kalman and the OLS models are used to forecast a fund's alpha. Starting from January 1970 until December 2002, the 1-factor and 4-factor OLS models (O1 and O4), and 1-factor and 4-factor Kalman models (K1 and K4) independently sort funds into 10 deciles based on their forecasted alphas. Decile 10 contains funds with the highest forecasted alphas. For each model, 10 equally-weighted portfolios are constructed from the stocks within the 10 deciles. For any fund to be included in any decile the following must be true: 1) the absolute value of forecasted alpha must be less than 2% per month, 2) the beta must be greater than 0 but less than 2, and 3) in the previous month the out of sample forecasted alpha and the difference between the realized excess return and the market return must have the same sign. Next, new decile portfolios $C_{i,j}$ are constructed by equally investing into funds that are selected by model i but not model j and that are selected by model j but not model i for each decile. Each portfolio is then held for 1 month and rebalanced at the beginning of the next month. If no disjoint funds are available for one specific decile during month t, then this decile invests in t-bills for that month. Also constructed are equally-weighted portfolios containing the intersection of the 20, 10 and 5 funds with the highest alpha forecasts by the two models. Finally, Panel A reports monthly excess returns and the Sharpe Ratios for these portfolios. Panel B reports the 4-factor (MKT, SMB, HML, and MOM) adjusted monthly returns.

A.	Monthly Excess Return				Monthly Sharpe Ratio			
Decile	$C_{O1,K1}$	$C_{O4,K4}$	$C_{O4,K1}$	$C_{O1,K4}$	$C_{O1,K1}$	$C_{O4,K4}$	$C_{O4,K1}$	$C_{O1,K4}$
1	0.0014	0.0020	0.0013	0.0016	0.0266	0.0410	0.0265	0.0323
2	0.0020	0.0022	0.0021	0.0022	0.0401	0.0455	0.0434	0.0451
3	0.0022	0.0025	0.0025	0.0022	0.0464	0.0538	0.0523	0.0460
4	0.0031	0.0027	0.0031	0.0027	0.0677	0.0593	0.0678	0.0606
5	0.0035	0.0032	0.0032	0.0035	0.0809	0.0743	0.0725	0.0815
6	0.0042	0.0035	0.0037	0.0038	0.0976	0.0828	0.0865	0.0891
7	0.0051	0.0044	0.0047	0.0049	0.1192	0.1020	0.1089	0.1151
8	0.0052	0.0053	0.0050	0.0051	0.1203	0.1210	0.1148	0.1169
9	0.0060	0.0057	0.0061	0.0057	0.1325	0.1252	0.1345	0.1277
10	0.0067	0.0061	0.0074	0.0057	0.1416	0.1279	0.1550	0.1234
Top 20	0.0070	0.0061	0.0075	0.0058	0.1437	0.1199	0.1485	0.1208
Top 10	0.0075	0.0067	0.0082	0.0062	0.1501	0.1260	0.1579	0.1241
Top 5	0.0075	0.0064	0.0081	0.0066	0.1480	0.1143	0.1494	0.1290
B.	4-factor adjusted return				T-ratio for 4-factor adjusted return			
1	-0.0023	-0.0023	-0.0028	-0.0024	(-2.74)	(-2.90)	(-3.72)	(-3.19)
2	-0.0018	-0.0020	-0.0020	-0.0017	(-2.72)	(-3.30)	(-3.25)	(-2.64)
3	-0.0017	-0.0018	-0.0016	-0.0019	(-2.86)	(-3.37)	(-3.15)	(-3.60)
4	-0.0009	-0.0016	-0.0010	-0.0015	(-1.84)	(-3.10)	(-1.97)	(-3.19)
5	-0.0006	-0.0008	-0.0009	-0.0006	(-1.24)	(-1.77)	(-1.96)	(-1.35)
6	-0.0002	-0.0008	-0.0007	-0.0005	(-0.47)	(-1.72)	(-1.52)	(-1.02)
7	0.0006	0.0000	0.0003	0.0004	(1.16)	(0.05)	(0.73)	(0.88)
8	0.0006	0.0009	0.0005	0.0005	(1.03)	(1.69)	(0.95)	(0.92)
9	0.0011	0.0014	0.0015	0.0012	(1.69)	(2.57)	(2.48)	(2.09)
10	0.0018	0.0020	0.0031	0.0011	(2.23)	(2.64)	(3.86)	(1.57)
Top 20	0.0023	0.0027	0.0036	0.0015	(2.54)	(2.94)	(3.97)	(1.90)
Top 10	0.0026	0.0036	0.0046	0.0018	(2.64)	(3.23)	(4.35)	(1.90)
Top 5	0.0025	0.0027	0.0040	0.0020	(2.09)	(2.10)	(3.21)	(1.74)

Table 13: **Out of Sample Performance: 12-month Holding Period.** For all domestic equity mutual funds having at least 3 years or 5 years of monthly return data (depending on the length of the training period), the OLS models are used to forecast a fund's alpha. Starting from January 1970 until December 2002, the 1-factor and 4-factor OLS models, OLS 1 and OLS 4 respectively, independently sort funds into 10 deciles based on forecasted alphas, which are estimated over the prior 3 or 5 years. Decile 10 contains funds with the highest forecasted alphas. For each model, 10 equally-weighted portfolios are constructed from stocks within the 10 deciles. Each portfolio is held for 12 months and rebalanced at the beginning of the next year. Finally, the table also constructs equally-weighted portfolios containing the 20, 10, and 5 funds with the highest alpha forecast by each model. Panel A reports the 4-factor (MKT, SMB, HML, and MOM) adjusted monthly returns realized during the entire period and the corresponding T-statistics. In Panel B, the portfolio returns are risk-adjusted by both four factors and one-month-lagged 4-factor returns (the 8 factor model).

Decile	OLS 1	OLS 2	OLS 1	OLS 4	OLS 1	OLS 4	OLS 1	OLS 4
	3 Years		5 Years		3 Years		5 Years	
	A1. Four Factor Monthly Alpha				A2. T-ratio			
1	-0.0011	-0.0013	-0.0006	-0.0019	(-1.26)	(-1.72)	(-0.83)	(-2.64)
2	-0.0004	-0.0011	-0.0001	-0.0007	(-0.70)	(-2.49)	(-0.23)	(-1.63)
3	-0.0002	-0.0007	-0.0001	-0.0003	(-0.29)	(-1.78)	(-0.25)	(-0.84)
4	-0.0004	-0.0007	0.0003	-0.0005	(-1.06)	(-2.03)	(0.65)	(-1.18)
5	-0.0004	-0.0006	-0.0002	-0.0004	(-1.12)	(-1.64)	(-0.45)	(-1.17)
6	-0.0005	-0.0004	-0.0004	-0.0002	(-1.57)	(-1.05)	(-1.04)	(-0.45)
7	-0.0003	-0.0002	-0.0004	-0.0001	(-1.00)	(-0.52)	(-0.92)	(-0.16)
8	-0.0002	-0.0001	-0.0005	0.0002	(-0.49)	(-0.31)	(-0.94)	(0.43)
9	-0.0004	0.0003	0.0002	0.0007	(-0.73)	(0.69)	(0.39)	(1.37)
10	-0.0003	0.0004	-0.0002	0.0012	(-0.33)	(0.52)	(-0.29)	(2.05)
Top 20	-0.0003	0.0008	-0.0004	0.0019	(-0.20)	(0.61)	(-0.40)	(2.48)
Top 10	-0.0001	0.0006	-0.0002	0.0025	(-0.04)	(0.38)	(-0.16)	(2.14)
Top 5	-0.0006	0.0009	0.0000	0.0028	(-0.29)	(0.44)	(0.02)	(1.60)
	B1. Eight Factor Monthly Alpha				B2. T-ratio			
1	-0.0011	-0.0012	-0.0009	-0.0019	(-1.14)	(-1.54)	(-1.07)	(-2.50)
2	-0.0002	-0.0009	-0.0001	-0.0008	(-0.24)	(-1.93)	(-0.13)	(-1.59)
3	-0.0002	-0.0006	-0.0001	-0.0002	(-0.29)	(-1.33)	(-0.18)	(-0.60)
4	-0.0003	-0.0007	0.0004	-0.0005	(-0.78)	(-1.75)	(0.98)	(-1.08)
5	-0.0004	-0.0005	-0.0003	-0.0004	(-1.01)	(-1.29)	(-0.61)	(-0.99)
6	-0.0005	-0.0004	-0.0003	-0.0001	(-1.51)	(-1.05)	(-0.84)	(-0.29)
7	-0.0003	-0.0001	-0.0005	-0.0002	(-0.89)	(-0.30)	(-1.00)	(-0.49)
8	-0.0001	-0.0000	-0.0004	0.0003	(-0.12)	(-0.04)	(-0.75)	(0.62)
9	-0.0002	0.0005	0.0002	0.0007	(-0.30)	(0.91)	(0.37)	(1.33)
10	0.0000	0.0006	-0.0003	0.0010	(0.02)	(0.83)	(-0.36)	(1.57)
Top 20	0.0001	0.0012	-0.0005	0.0016	(0.06)	(0.84)	(-0.45)	(1.95)
Top 10	0.0005	0.0011	-0.0005	0.0019	(0.28)	(0.61)	(-0.36)	(1.55)
Top 5	0.0001	0.0013	-0.0002	0.0019	(0.03)	(0.60)	(-0.08)	(1.04)

Table 14: **Fraction of Funds with Missing Data: Year by Year.** This table displays the average annual percentage of funds that have missing data within a five year window prior to the month in question. In each month, the number of funds that were founded at least five years prior and were still in business but for which data over that interval is missing constitutes the numerator. The denominator is the total number of funds that were founded at least five years prior to the month in question and were still in business. This number is then averaged for the 12 months within each year and multiplied by 100.

Percentage of funds with missing data in each year.

Year	Percent	Year	Percent	Year	Percent	Year	Percent
1970	8.69	1980	0.25	1990	0.47	2000	0.12
1971	8.39	1981	0.32	1991	0.21	2001	0.14
1972	4.08	1982	0.49	1992	0.00	2002	0.13
1973	0.67	1983	0.30	1993	0.01		
1974	0.80	1984	0.28	1994	0.24		
1975	0.36	1985	0.00	1995	0.32		
1976	0.62	1986	0.16	1996	0.30		
1977	1.00	1987	0.38	1997	0.27		
1978	1.05	1988	0.33	1998	0.16		
1979	0.34	1989	0.40	1999	0.13		

Table 15: **Return History for Funds with Omitted Data.** This table examines the returns for the 39 funds in the CRSP database that have missing data and more than 30 months of return data (after excluding missing data) in 1970-2002 period. Panel A looks at their overall returns. "Ret-MktRet" is the excess return over the market for all available dates from 1970-2002. The "CAPM Alpha" and "Carhart Alpha" are the alphas from the one and four factor regressions that use all available return data from 1970-2002. Panel B compares the returns on these funds during training periods with those they generate during holding periods. The row t-60:t-1 Alpha displays the average training period alpha from the 60 month testing period. The t+1:t+12 Ret-MKT lists the difference between the fund's return during the holding period and that of the market. The AlphaDiff row is simply the difference between the t+1:t+12 Ret-MKT and t-60:t-1 Alpha rows. Columns under the labels "Pass" only include funds that pass a period t-1 back test. The "Pass ratio" equals the fraction of funds that pass the back test. The total number of pooled training periods is 1355.

A. Raw Above Market Returns and Regression Alphas.

	Ret-MktRet	CAPM Alpha	Carhart Alpha
Mean	-0.0109	-0.0025	-0.0005
Std. Dev	0.0137	0.0078	0.0080
Std. Error	0.0022	0.0013	0.0013
T ratio	-4.9971	-1.9655	-0.3801

B. Training Period and Holding Period Returns and Alphas.

	CAPM Alpha	CAPM Alpha (Pass)	Carhart Alpha	Carhart Alpha (Pass)
t-60:t-1 Alpha	-0.0007	-0.0013	0.0009	0.0003
T ratio	-2.6385	-4.6901	2.7580	0.8741
Pass ratio	0.4849		0.5114	
t+1:t+12 Ret-MKT	-0.0005	0.0003	-0.0005	-0.0000
T ratio	-0.8504	0.3742	-0.8504	-0.0209
AlphaDiff	0.0003	0.0016	-0.0013	-0.0003
T ratio	0.4530	1.7458	-2.1072	-0.3291

Table 16: **Out of Sample Performance (With Parameter Forecast Filters, With Back Testing, 1985-2002).** This table reproduces the analysis in Table 5 Panel B except that the holding period data begins in 1985.

Decile	OLS 1	OLS 4	Kal 1	Kal 4	OLS 1	OLS 4	Kal 1	Kal 4
	B1. Four Factor Monthly Alpha				B2. T-ratio			
1	-0.0040	-0.0039	-0.0033	-0.0042	(-2.98)	(-2.58)	(-2.56)	(-3.50)
2	-0.0025	-0.0033	-0.0023	-0.0020	(-2.32)	(-3.73)	(-2.31)	(-2.34)
3	-0.0029	-0.0025	-0.0019	-0.0019	(-3.26)	(-3.46)	(-2.36)	(-2.57)
4	-0.0016	-0.0021	-0.0015	-0.0022	(-2.00)	(-2.77)	(-2.19)	(-3.39)
5	-0.0014	-0.0012	-0.0018	-0.0011	(-2.35)	(-1.78)	(-3.10)	(-1.70)
6	-0.0006	-0.0014	-0.0012	-0.0011	(-1.04)	(-2.14)	(-2.34)	(-1.97)
7	-0.0003	-0.0004	0.0002	0.0005	(-0.45)	(-0.56)	(0.36)	(0.79)
8	0.0000	0.0001	0.0002	0.0002	(0.03)	(0.12)	(0.33)	(0.39)
9	0.0007	0.0010	0.0009	0.0012	(0.91)	(1.27)	(1.30)	(1.67)
10	0.0011	0.0032	0.0021	0.0016	(0.84)	(2.48)	(1.98)	(1.58)
Top 20	0.0019	0.0047	0.0032	0.0015	(1.25)	(2.51)	(2.32)	(1.14)
Top 10	0.0028	0.0060	0.0030	0.0023	(1.44)	(2.68)	(1.97)	(1.39)
Top 5	0.0026	0.0045	0.0030	0.0009	(1.15)	(1.84)	(1.64)	(0.49)