

# **Evaluation and Elimination in Mutual Fund Selection**

Robert L. Padgett, CFA, CIMA  
Ted A. Ponko, CFA

Klein Decisions  
68 T.W. Alexander Drive  
Research Triangle Park, NC 27709

919.233.6767

# Evaluation and Elimination in Mutual Fund Selection

## *Executive Summary*

Financial planners need an effective means of analyzing and selecting mutual funds. Many rely on traditional screens that basically function as pass/fail tests. Funds meeting all criteria move on for further evaluation while those that fail are removed from consideration. Although easy to apply, the screening process does not typically provide comprehensive, actionable results. Rather than comparing funds, it simply produces an unordered list of survivors. The planner must invest additional time and effort in additional screening and sorting before arriving at a final answer.

Factor models offer a superior alternative. Like traditional screening, they start with the selection of criteria for evaluation (the “factors”), but then allow the planner to weight the factors based on their relative importance. Funds receive a single score based on the sum of their performance on all the weighted factors. The result is a rank ordering with funds that have the highest scores rising to the top. Unlike the screen’s list of survivors, these funds don’t simply meet a minimal standard; *they have the highest combination of all the weighted factors*. The planner can confidently begin the final stage of due diligence starting with the top fund and working down without all the additional screening and sorting. Enterprising planners with a data source and spreadsheet skills can create their own factor models, while others may choose to utilize the commercially available alternatives.

Driven by the explosive growth of 401(k)s and individual investing, mutual funds have arguably become the nation's most favored investment vehicle. The number of funds has increased just as dramatically. Today there are over 8000 unique funds,<sup>1</sup> or nearly three times that many considering all share classes. This offers planners an opportunity to distinguish themselves by the funds and asset allocations they use in building client portfolios. But with this opportunity comes a challenge: Planners need an evaluation process that considers all appropriate alternatives, renders conclusive and actionable results, and is consistent and repeatable for selection and ongoing monitoring.

Some planners devise their own analysis tools using their time, expertise, a data source, and spreadsheet software. Others use fund analysis tools readily available from a wide array of commercial sources. Understandably, the quality of these tools, the investment in time required, and the value of the output varies tremendously. The underlying methodology varies also. While some tools employ linear screening models others have advanced to a multi-factor analysis. An understanding of the differences between these two approaches is essential when deciding the process and tools used in fund evaluation and selection.

### ***Traditional Screening***

Despite their variety, most fund analysis strategies can be divided into two categories. The first is essentially a process of elimination based on pass/fail tests. Each test is applied independently and only funds that survive the entire series move on to more qualitative analysis. Those that make it through this stage can then be used as the basis for the recommended list.

This traditional screening process is easily understood and implemented. The planner selects the screening criteria (e.g., 3-year annualized total return, expense ratio, manager tenure, etc.) and the universe to examine (e.g. domestic large cap equity funds), and the software does the rest. Tools can be found at many financial websites and more sophisticated versions are available through providers such as Morningstar and Value Line. Planners seeking more depth and flexibility tailor their screens by importing fund data into spreadsheets and then sorting based on their preferred criteria. Funds that make their way to the top of the list for each sort can then move on to a more qualitative "due diligence" step.

Planners with limited time and staff appreciate the simplicity of the screening process. By quickly narrowing the universe of potential funds to a more manageable level, it allows them to focus their limited time and resources on the due diligence phase. Arguably, this is screening's greatest strength. In addition, the process is simple to perform and is easily understood by clients, both pluses for the planner.

As an example, consider a screen of large cap blend funds using October 31, 2007 data. To keep things simple, we used the five most popular equity screening elements used by advisors as reported in a 2006 survey conducted by the Financial Research Corporation

on behalf of the FPA<sup>2</sup>: Expense Ratio, 5-Year Performance, Manager Tenure, Risk-Adjusted Performance, and 3-Year Performance.<sup>3</sup> Because we are evaluating large cap equity funds, we set the limits for each screen slightly over the corresponding values of the Large Blend *Morningstar* category average. The presumption – like that used in many screening processes – is that superior actively managed funds should outperform their peers. The screens are summarized in Table 1 which lists them in their order of popularity.

After applying these screens in Morningstar *Principia*'s screening software, the 1,951 funds in the category were reduced to a list of 134. In this case, only 6.9% of the Large Blend funds passed all five screens, clearly demonstrating the processes' ability to narrow the universe of alternatives. The screens were easily set up and run, and the results were quickly returned, all of which would be appreciated by time-constrained planners.

Screen	Requirement
Expense Ratio	≤ 1.25%
5-Year Total Return NAV	≥ 13.65%
Average Manager Tenure	≥ 3 Years
5-Year Sharpe Ratio	≥ 1.08
3-Year Total Return NAV	≥ 13.34%

Unfortunately, traditional screening processes have inherent flaws as well. They stem from three primary sources:

1. *Funds are removed from consideration even if they fail the least important criterion by a minimal amount.*

There are no degrees of success or failure in a pass/fail screen. A fund either passes or fails. A fund could pass all others by wide margins, but if it fails only one by a small amount, it is still eliminated. Table 2 shows five funds that failed the screens of Table 1.

Fund	Expense Ratio ≤ 1.25%	5-Yr TR NAV ≥ 13.65%	Avg Mgr Tenure ≥ 3 Years	5-Yr Sharpe Ratio ≥ 1.08	3-Yr TR NAV ≥ 13.34%
Tocqueville (TOCQX)	<b>1.30</b>	19.98	15.8	1.35	17.73
New Century Capital (NCCPX)	<b>1.27</b>	15.43	12.7	1.35	15.62
American Century Equity Growth (AMEIX)	0.47	14.72	8.2	1.19	<b>13.29</b>
Goldman Sachs Structured Tax-Managed Eq (GCTIX)	0.69	15.01	4.7	1.20	<b>13.33</b>
Neuberger Berman Guardian (NGUAX)	0.88	16.27	4.2	1.25	<b>13.33</b>

The first two have expense ratios that are 0.05% or less over the maximum. This is enough to eliminate them from consideration even though their results on the remaining screens are substantially over the respective limits. The last three funds have 3-year total returns that fall from .01% to .05% below the required minimum and again, despite their encouraging scores on the other screens, each is eliminated. In all these instances, otherwise acceptable funds are eliminated for missing one test by a marginal amount. This is an unavoidable consequence because screens have no means of distinguishing the magnitude of failure.<sup>4</sup>

2. *All criteria are equally weighted.*

Basic screening processes are simple pass/fail tests. Since failure to satisfy any one of the criteria is sufficient to remove a fund from consideration, each factor carries the same weight. That is why the final two funds in Table 2 are eliminated despite only failing the least important screen by .01%. Even if the planner feels that one factor is substantially more important than another, there is no way to reflect this difference in a simple pass/fail screen. As a result, this encumbers the planner's ability to incorporate their expertise in the process.

3. *Funds can only be evaluated on one criterion at a time rather than simultaneously.*

By its very nature, screening is a process of elimination. It is a pass/fail test against each criterion individually rather than the group as a whole. Again to return to the example above, Tocqueville has superior results in regard to the final four criteria yet it still fails the overall screen by falling just short (.05%) in the first test. Since each test must be passed independently, superior performance in regard to the other four has no bearing on passing the fifth.

As a related consequence, just as the screen offers no means of distinguishing between funds that fail, it also offers no means to differentiate among those that pass. Although the process measures funds against various criteria, it does not measure them against *one another*. In the example, the universe is reduced to 134 funds but the work is not yet done. The planner must still find a way to select the best of the lot. While it is possible to tighten the screening criteria to produce a smaller list of survivors, the job still is not complete until it yields only a few "top" funds, thus exacerbating the problem described in item 1, elimination of excellent funds on insignificant screening margins. On the other hand, it may be determined that surviving funds that pass the various tests by greater magnitudes will be more likely to be added to recommended list, yet this is an additional *qualitative* judgment and not a result of the quantitative screen itself. In any event, the planner faces more work before arriving at the final answer.

Screening is like having an 18-hole golf tournament where anyone who makes a bogey is immediately eliminated. When it is over you only know that those who made it through all eighteen holes did not have a bogey but you do not know who had the best round. You cannot tell the one with ten birdies from the one who simply shot even par. In fact, you do not know if someone eliminated on the eighteenth hole had actually birdied the previous seventeen. Like screening, this tournament is based on eighteen separate individually applied tests rather than one overall score.

That is why golf tournaments are not run this way. When you consider the total score, you not only know who had the best round, you know the relative finish of everyone else, too. Weighted factor models, the second approach to fund analysis, yield comprehensive, ordered results while avoiding all of screening's basic weaknesses.

### ***Weighted Factor Models***

Like traditional screening, factor models allow funds to be measured against various criteria. Unlike traditional screening, however, criteria are weighted based on their relative importance. Funds are then measured against these weighted factors simultaneously, resulting in a single score. Because this is not a process of elimination, all funds receive a score and can be rank ordered against their peers, just like the total scores in a golf tournament.

In its most basic sense, a weighted multi-factor model takes the following form:

$$Score = \sum_i^n \beta_i f_i + \beta_2 f_2 + \dots + \beta_n f_n$$

Where:

*Score* = overall score for ranking purposes

$\beta_i$  = weighting of factor *i*

$f_i$  = specific factor score

*n* = number of factors in the model

Factor weights are usually expressed as percentages which typically total to 100% (except in cases where leverage and/or short-selling are allowed):

$$\sum_i^n \beta_1 + \beta_2 + \dots + \beta_n = 100\%$$

As with traditional screening, the planner selects the factors to be used in the analysis. Each fund's actual performance is measured relative to each factor, and this result is then adjusted by the respective factor weight. Finally, values for all factors are summed to yield a single overall score for each fund.

Rather than simply producing a list of funds that "passed" the analysis, the factor model provides a means of rank ordering *all* the funds. This not only reveals how each fared but also measures them in relation to one another. Funds selected to move on to the next step of evaluation can be drawn from the top of the list. Since there is no "pass" or "fail", the planner can decide the appropriate cutoff.

An enterprising planner can create the factor model with spreadsheet software, imported fund data, and considerable time. Macros can be written to measure each fund on each factor and assign the appropriate scores. These totals can then be summed to come up with a unique score for each fund. Finally, funds can be sorted based on their overall scores to create a rank ordering. Funds at the top can then move to the next due diligence step. Specially designed software can save both time and effort and is available to simplify this process.

### ***Scaling***

The most challenging tasks in the creation of a factor model are the assignment of point totals for fund results and the factor weighting procedure. For point totals, the actual range of each factor can be used to assign scale-based scoring. For example, funds that

have the highest value for the factor receive 100 points while those with the lowest value get 0.<sup>5</sup> All other funds within the measurement universe would fall between these two extremes and receive factor scores in proportion to their position within the actual range.

As an illustration, consider the 5-Year Total Return NAV from the prior example. Although the screen minimum was set at 13.65%, funds in the category actually ranged from +3.64% to +35.13%. Tocqueville's 5-year total return was 19.98% which is 51.88% of the distance between the extremes, so it receives 51.88 points for this factor. New Century Capital's 5-year total return of 15.43% is 37.44% along the way between the extremes giving it 37.44 points. The other funds' scores are calculated in a similar manner.

### *Setting Factor Weights*

Part of the benefit of using a weighted factor model is the user's ability to assign different weights to each factor to reflect relative importance. This is a much more intuitive approach than equal-weighted screening for the planner who believes some factors are more important than others. The trick is in quantifying these beliefs.

There are a number of ways to assign weights ranging from the simple to the complex. Those falling into the latter category attempt to gauge each factor's contribution through a statistical analysis of past performance. Results are then translated into weights for each factor. The cost and time of this procedure is likely to be beyond the reach of many planners.

At the other end of the spectrum lie some very simplistic modeling tools. These allow the planner to select a number of factors and then assign weights to each. The tools then apply the model to the user's universe of funds, calculate the scores for each, and display the ordered results. Applets for this approach are available at several popular investing websites. Alternatively, with a little work a planner with access to a database of fund information (such as Morningstar's *Principia*) can create and weight a simple factor model in a spreadsheet program.

There are, however, important limitations to this process, the most important being the lack of rigor. While the planner may have a sense of which factors to include and their relative importance, the simplest tools require the planner to heuristically translate those feelings into numerical weights. To simplify the process, some require increments of 5% or 10%. A planner creating his or her own model on a spreadsheet will likely use similar shortcuts. Without further assistance, it is difficult – if not impossible – for the planner to reasonably determine the appropriate values with such precision. In essence, the simplest tools lack the analytical basis necessary to generate reliable results.

A better solution can be found in the next generation of modeling tools that incorporate a systematic weighting process. An integral part of these tools is an analytical process to quantify the user's preferences into appropriate weights, generally using either adaptive conjoint analysis (ACA)<sup>6</sup> or the analytic hierarchy process (AHP).<sup>7</sup> Studies have shown that while humans can usually distinguish the relative importance between two factors,

they have much less confidence when they are called upon to order more than two. These tools break the decision process down into a series of paired choices, never requiring the user to gauge the relative weight of more than two alternatives at time. The responses can then be used to consistently calculate the specific factor weights. The process then follows the path of the simple tools resulting in a fund ranking, but now based upon rigorously created factor weights.

We used Klein Decisions' *K<sup>4</sup> Fund Selection* to evaluate the same universe of funds screened earlier. This tool gives the user the choice of creating an equal-weighted model, assigning his or her own weights to each factor, or to apply (ACA) by answering a short series of importance and trade-off questions to determine the appropriate weights. We used the second option to allow the weights to reflect the usage levels from the FPA survey. As with the earlier screen, the goal was not to create the "best" model, but rather one that reflects what planners are doing.<sup>8</sup>

### ***Comparing Results***

Although there are a number of differences between the two sets of results, two jump out immediately. First, the screen returns only 134 funds while the factor model includes all 1,521. That may appear to be a plus for screening until you consider the other difference: The factor model's results are rank ordered, the screen's are not. The factor model includes all funds in the analysis and rank orders them so that the top funds can be identified assuring that the best funds are not missed and that a definitive measurement is objectively determined.

By its very nature, the screening process is a means of narrowing the universe of funds to a manageable number that can then be further culled and evaluated. In this case, the screen leaves 134 funds for further consideration. As discussed above, the planner still has additional work to do. The screens can be rerun with tighter criteria or additional screens can be added to further limit the number of survivors. This is also the point at which many planners run various sorts on fund characteristics using spreadsheet software. The hope is to find funds that rise to the top on several sorts, suggesting they may be the most appropriate for further consideration. Either way, more work has to be done before the final analysis and selection can take place, and whatever form it takes, it will demand more of the planner's time.

On the other hand, the factor model has already ranked all the funds, in this case from 1 through 1,951. The planner using this approach can immediately begin the final stage of due diligence starting with the top fund and working down the list as far as deemed necessary. The total size of the ranked universe really does not matter because the fund selection process will rarely go beyond the top funds. The reason stems from the way the funds are ranked.

Unlike screening which tests each fund on one criteria at a time, a factor model scores them on all criteria *simultaneously*. Funds that have the highest weighted scores on all factors rise to the top. Unlike the screen's list of survivors, these funds don't simply meet a minimal standard, *they have the highest combination of all the weighted factors*. In

short, they are the best fit for the planner’s preferences. The planner might find them in the screen results, but only after a considerable amount of additional work. In the process, he or she will also miss some good alternatives.

Consider the funds on Table 3 drawn from the Large Blend universe of the earlier example. In addition to the results for the five factors from Table 2, its two final columns show results against the screens and factor model ranking. The first three are the top rated funds from the factor model. As you would expect, all three easily pass the screens, too, with returns well above the minimums and expense ratios comfortably under 1.25%. Given that, one might wonder if one approach has the advantage over the other, but the major difference doesn’t appear on the chart: *The factor model automatically brings these fund to the top, screens do not.*

TABLE 3							
Fund	Expense Ratio ≤ 1.25%	5-Yr TR ≥ 13.65%	Avg Mgr Tenure ≥ 3 Years	5-Yr Sharpe Ratio ≥ 1.08	3-Yr TR ≥ 13.34%	Pass All Screens?	Factor Model Ranking
CGM Focus (CGMFX)	1.02	35.14	10.2	1.54	40.02	Yes	1
American Funds Fundamental Investors (RFNFX)	0.35	19.62	14.0	1.58	19.63	Yes	2
Hartford Capital Appreciation (HCAYX)	0.73	22.48	10.1	1.62	21.91	Yes	3
Tocqueville (TOCQX)	1.30	19.98	15.8	1.35	17.73	No	31
New Century Capital (NCCPX)	1.27	15.43	12.7	1.35	15.62	No	99
BNY Hamilton Large Cap Equity (BNEIX)	1.03	13.65	4.0	1.31	14.93	Yes	477
New Covenant Growth (NCGFX)	1.08	13.90	5.0	1.12	13.91	Yes	624

In addition, the factor model presents all viable alternatives including those excluded by the screening process. Again consider Tocqueville and New Century Capital which were screened out because their expense ratios slightly exceed 1.25%. Despite that, their returns and Sharpe Ratios are considerably higher than the screen minimums, allowing the factor model them to rank 31 and 99, respectively. With the exception of their expense ratios, they dominate the final two funds (which did survive all screens) in every other category. To put this in context, 70 of the 134 funds passing the screen have a factor model score over 99. This indicates that despite missing one screen by a small amount, Tocqueville and New Century Capital still score better on the combined weighted factors than over half the funds passing all screens. This is a good example of how funds that score under the screening limit versus one criteria may still fare well in the overall rankings if they excel against all other factors. The screening process cannot measure funds against all factors simultaneously while factor models are designed to do just that. When investors buy a fund, they buy all of its characteristics together, so isn’t this the way funds should be evaluated? .

Of course the planner may believe some features or levels of performance are absolutely essential. Alternatively, the client or investment policy statement may require all funds have (or not have) specific characteristics. Without meeting these requirements, the fund is unacceptable. This is where screening adds value, yet it is not incompatible with factor models, either. The *K<sup>4</sup> Fund Selection* tool allowed us to apply screens to the ordered list to eliminate funds lacking “essential” features. As suggested above, applying all five of the screens may eliminate some valuable alternatives, yet for the purpose of comparison we applied the five with the limits shown in Table 3. Not surprisingly, the factor model’s results were reduced to the screen’s 134. The big difference, however, was that *they were still rank ordered* as opposed to the screen’s collection of survivors. Not only that, by rank ordering the funds first and then screening for essential features, it is possible to see which funds are being eliminated and where they originally ranked. A screen alone will not do that.

### ***Conclusions***

After an examination of screening vs. factor models for fund evaluation, we can conclude that factor models offer a superior methodology for fund evaluation based on three things:

1. All funds under consideration are rank ordered ensuring that the top funds across all criteria are identified and that superior funds are not missed by the evaluation process.
2. Criteria for evaluation of funds can be weighted ensuring that their relative importance is considered in the evaluation process.
3. A comprehensive actionable result is produced at the end of the process requiring no additional time-consuming steps to actually find the top funds. .

Even though factor models do provide a superior evaluation methodology, cynics may fear the factor model’s rigor adds little value other than the potential to dazzle uninformed clients with a highly quantitative process. After all, a factor model will yield a rank order of funds even if the model itself is poorly constructed. The mere fact that it is more rigorous than a simple screen does not necessarily lead to better results.

There is little reason to believe the results of any poorly constructed evaluation process will lead to superior funds and factor models are no exception. However, diligent planners who are willing to put in the thought and time to create a thorough fund evaluation process will find that factor models are more effective than screens in locating funds with the highest levels of desired properties. The goal is not to find an impressive process but rather one that efficiently locates the most appropriate funds.

Even so, some may question if the use of a factor model will really lead to superior results. While they acknowledge it is easy to see which funds have the highest scores in a factor model, they question if the results are truly more predictive than those of a traditional screening process.

When planners evaluate funds, they seek those that they believe will have the greatest likelihood of meeting their objectives in the future. It is not difficult to see which funds performed well on certain factors in the past, but investments can only be made prospectively, not retroactively. As a result, most fund evaluation focuses on finding funds with those characteristics believed to be most predictive of future performance. A great deal of research has been conducted in this regard, yet no factors have proven to be definitively superior predictors. Some planners may focus on alpha, others on consistency, and still others on downside protection. As the survey cited earlier suggested, planners do focus on some historical factors more than others.

Although selecting funds that will perform well in the future is an important goal for planners and investors, neither screens nor factor models measure the predictive value of specific factors. Instead, they evaluate *funds* and the degree to which they display the various factors believed to be important based on the planner's expertise. This is where factor models excel.

Traditional screens test funds on one factor at a time to yield an unordered collection meeting minimum standards. Factor models evaluate all funds on every factor, to generate a combined score reflecting the total levels of all factors. Funds with the highest scores have the greatest levels of the planner's desired factors. These factors may or may not be predictive of future performance, *but they are the ones the planner sought*. Both fund evaluation processes provide a way to find these funds, but because of the combined scoring process and ability to rank order funds, factor models are more efficient.

Finally, some may wonder if there are specific circumstances where screening may be preferable to the use of a factor model. Could screening be more appropriate for passively managed funds, or bond funds, or international funds? How is a planner to decide?

These questions bring us back to the fundamental differences between the two processes. Factor models allow the planner to rank order a particular universe of funds based on the importance of each factor and their total weight-adjusted level for each fund. Funds are scored and not eliminated. Funds with greater degrees of the desired factors score higher than those with less. While the planner may use one set of factors in a particular sector of the market and a different one in another, the basic process works in all. Efficiently finding funds with the highest degree of desired characteristics should be the goal in all areas of the market.

But screening has a role, too. For example, if a fund has high marks on all factors yet the manager has only been at the helm for two months, the fund's score loses some (if not all) of its significance. Applying a manager tenure screen to the ordered list of funds can alleviate this problem by eliminating funds with new management. As discussed above, screens may also be used to flag funds lacking other "must-have" features.

The use of screens and factor models should not be viewed as mutually exclusive alternatives but rather as complementary parts of the fund evaluation process. Factor

models score funds based on the degree to which they display the factors sought by the planner. Screens can then eliminate those funds failing to have the appropriate level of specific essential features. The result is a rank ordered list of funds with the highest levels of the planner's desired characteristics and also the planner's specific investment requirements. No additional sorts or iterations are needed. This is the final answer to the quantitative part of the evaluation.

Even the most sophisticated quantitative model is no substitute for the planner's knowledge, experience, and investment expertise. These must still come into play in the qualitative interpretation of results and the final fund selection. The best models are those that enable the planner to incorporate his or her knowledge and skills from the outset. The challenge is to find a complete, concise, actionable, and repeatable way to do this. Weighted factor models are a solution worth considering.

---

<sup>1</sup> *Investment Company Institute Fact Book*, 47<sup>th</sup> edition, p.93.

<sup>2</sup> "How Independent Advisors Select Independent Products", *Journal of Financial Planning*, April 2006, Article 9.

<sup>3</sup> The authors are not making the argument that these are the best criteria; in fact we suspect they are not. However, the focus here is not on the predictive quality of the screen, only the methodology. For insight into the former, please see William G. Droms' compendium of research, "Hot Hands, Cold Hands: Does Past Performance Predict Future Returns?", *Journal of Financial Planning*, May 2006, Article 7.

<sup>4</sup> A related issue occurs when a fund fails or is unable to provide data for a particular criterion. Missing data is a reality among all investment databases. The data may be absent because the fund has yet to reach the required track record, fund personnel erred in data submission, etc. Screens treat missing data as failures. As a result, missing data for any criteria will eliminate a fund from consideration.

<sup>5</sup> In situations where there are significant outliers at either or both extremes, it may be helpful to set the endpoints closer to the main body of the distribution. This will result in assigning maximum and/or minimum values to a range rather than a single value.

<sup>6</sup> Conjoint analysis has been successfully used in the field of market research since the late 1970s and has been a mainstream technique for assessing consumer preferences for over 15 years. Adaptive Conjoint Analysis (ACA) asks respondents two types of questions: Importance and trade-offs. The first asks the user to rate the importance of the difference between a positive outcome and a negative outcome for each attribute in the evaluation. Based on these rankings, pairs of attributes are then presented to assess the respondent's reaction to compromise. By analyzing these responses, it is possible to determine the value the respondent places on each specific attribute. In the case of fund evaluation, the attributes would be the fund characteristics deemed important by the planner.

<sup>7</sup> For a discussion of how AHP can be used in the investment decision process, see Hakan Saraoglu and Miranda Lam Detzler, "A Sensible Mutual Fund Selection Model", *Financial Analysts Journal*, May-June 2002, pp. 60-72.

---

<sup>8</sup> If the goal had been to create the best or optimal model, we would have relied upon the option to use adaptive conjoint analysis to assign weights to the various factors. Our goal here was to reflect planners' preferences captured in the FPA survey.