On luck versus skill when performance benchmarks are style-consistent

Sam Agyei-Ampomah[a], Andrew Clare[b,*], Andrew Mason[c], and Stephen Thomas[b]

[a] *Cranfield School of Management, Cranfield University, Cranfield MK43 0AL.*
[b] *The Sir John Cass Business School, City University, London, UK.*
[c] *Surrey Business School University of Surrey Guildford GU2 7XH.*

This Version: April 2015

Abstract
We firmly believe that style-appropriate, investible benchmarks not only provide a more parsimonious way of describing manager performance, but also that their use better aligns performance evaluation with the real world performance targets of fund managers'. It is against such benchmarks that managers should be judged. With this principle foremost in our approach, in this paper we use style-consistent benchmarks to determine whether any observed alpha produced by a sample of U.S. equity funds is due to skill or to luck. We find that different segments of the market, ranging from large-cap growth to small-cap value, exhibit different levels of skill and luck. Our results also show that the use of standard multi-factor models underestimates managerial ability and overstates the proportion of funds whose abnormal performance can be attributed to chance rather than to skill, when compared against the use of style-consistent practitioner benchmarks. We also find that a single factor performance evaluation model that uses Russell style indices consistent with the style orientation of a fund and market practice provides a parsimonious way of accounting for fund performance. Finally, our findings should be of particularly relevance in mutual fund markets where the risk factors commonly used in the academic literature to evaluate manager performance – SMB, B/M, MOM and others – are not readily available

JEL classification: G1
Keywords: Mutual fund performance; style benchmarks; skill versus luck

---

*Corresponding author. Tel.: +442070405169

E-mail addresses: sam.agyeiampomah@cranfield.ac.uk (Agyei-Ampomah); a.clare@city.ac.uk (Clare); andrew.mason@surrey.ac.uk (Mason); S.H.Thomas@city.ac.uk (Thomas).

1. Introduction

It is difficult to understate the importance of the US mutual fund industry; 92 million individuals, or 54 million households in the U.S. own mutual funds and these funds hold 24% of U.S. corporate equity. At the end of 2012 the $13 trillion of mutual fund assets were approximately the same size as the assets of all commercial banks in the United States. (Investment Company Institute, 2013, Federal Reserve, 2013). Of this $13 trillion, $4.3 trillion are invested in U.S. domestic equities, and 83% is managed on an active basis. Given its importance it is essential to strive to explain the existence or absence of skill amongst those charged with the oversight of these actively managed assets. However, there is a significant amount of evidence in the finance literature that suggests that actively managed mutual funds underperform the market and/or their assigned benchmarks on average (at least net of fees). For example, Lakonishok et al (1994) and Carhart (1997) (amongst many others) find little evidence of skill. Wermers (2000) finds that skill may exist at the gross of fee level but this does not filter through to the ultimate investor through net of fee returns. A number of researchers have found that performance persistence tends to exist amongst the poorest performers (see for example Goetzmann and Ibbotson (1994) or Berk and Tonks (2009)), although Grinblatt and Titman (1992), Goetzmann and Ibbotson (1994) and Fama and French (2010) document evidence of positive persistence for the very top performing funds.

Many of the studies that attempt to establish the existence, or otherwise of skill amongst active fund managers, apply standard multi-factor models as the relevant benchmarks for performance comparison, with the factors popularised by Fama-French serving as the major point of reference here. However, recent empirical studies have shown that alphas obtained from these standard multi-factor models can misstate managerial ability (see for example

Cremers et al (2012), Argon and Ferson (2006), or Angelidis et al (2013)). However, as well as potentially misstating the degree of skill, the use of multifactor benchmarks embodies the implicit suggestion that, in addition to the market portfolio, fund managers should invest in hedge portfolios that compensate for risks associated with small, growth and momentum stocks. However, the major drawback of this 'advice' from the academic community to the fund management community is that these hedge portfolios are not investible when one takes into account capacity constraints and transaction costs, particularly the transactions costs of shorting even the largest, most liquid stocks (see Huij and Verbeek (2009)). Christopherson et al (2009) provide an excellent description of the desirable characteristics of a financial market benchmark. A benchmark should provide a "naïve" representation of the set of investment opportunities facing investors; in our case a style group of mutual funds. The index should be investible and cover the practical opportunities for an investment style. It should be float-adjusted, that is, it should be based on the market capitalisation of tradable shares. Perhaps more importantly, the benchmarks should have a clear, simple and transparent construction methodology that can be easily replicated by others. The risk factors that comprise the multi-factor performance evaluation models are not replicable, investible benchmarks and as such their use in performance evaluation raises the question as to what exactly is being evaluated with their use.

With regard to the constraints that managers have to operate within by their mandates such constraints can vary significantly across different investment style groups. If, for example, a fund markets itself as a 'Large Cap Growth fund', then managers of the fund are constrained in terms of the stocks that they can hold in the fund's portfolio as a result of regulatory requirements and other restrictions from the fund's sponsors and trustees (see Clarke et al (2002)). Any measure of managerial performance that ignores such constraints will therefore be inefficient in assessing managerial ability. Indeed, Kothari and Warner (2001) and Angelidis et al (2013) argue that standard mutual fund performance measures are unable to identify significant abnormal performance if the fund's style characteristics differ from those of the benchmark portfolio, while Cremers et al (2012) highlight biases and shortcomings of the standard factor models. Chan et al (2009) show that for conventional size and value style U.S. funds over the period 1989-2001, there is disagreement about the *sign* of excess returns in approximately one quarter of cases, while absolute annual abnormal returns can also differ by large magnitudes depending on the choice of benchmark. Consistent with the predictions of Kothari and Warner (2001), Angelidis et al (2013) and Cremers et al (2012), we show in this paper that average performance of the different style groups using style-consistent benchmarks is economically different from those obtained using the standard multi-factor models, by as much as 0.34% per month in the case of small cap growth funds. All the small cap style groups (growth, blend and value), on aggregate, generate significant superior performance (net of cost) when measured against their respective style benchmarks. Using benchmarks that ignore the fund manager's mandated investment style and philosophy can therefore affect performance evaluation by misstating managerial skill. Evidence of differential performance across different style groups has also been reported by Chen, Jegadeesh and Wermers (2000) and Kosowski et al (2006).

Although using an appropriate, investable benchmark is a necessary component in the identification of manager skill, it is not however necessarily sufficient. Recent empirical studies have shown that the identification of significant positive alphas is not sufficient to confirm whether superior investment skill exists. These studies examine whether abnormal fund performance, where it is observed, is due to managerial skill or pure luck. Cuthbertson et al (2008), Fama and French (2010), Busse et al (2010) and Barras et al (2010), among others, all provide evidence to suggest that the limited amount of skill that can be identified is largely attributable to good luck rather than to skill. These studies however examine the issue of luck versus skill across the entire cross-section of funds as a whole and use standard multi-factor benchmark models, ignoring the potential differential performance across style groups and industry convention which emphasises peer groups and investable benchmarks. Kosowski et al (2006) find evidence of skill amongst a 'sizeable minority' of funds that cannot be attributed solely to luck. When they look at fund performance on the basis of a small number of prospectus objectives they find evidence of skill for growth-oriented funds but none for income-oriented funds.

Although a number of studies mentioned above show that alphas that are obtained from standard multi-factor models misstate managerial ability compared to industry style benchmarks, we are not aware of any study that provides a similar comparison with regard to the 'luck versus skill' debate. If the alphas generated by multifactor models can be economically different from comparative alphas from the style-consistent benchmarks, then luck versus skill analysis based on the multifactor models also has the potential to be misleading. This observation is important because evidence of skill using the style-consistent benchmarks can provide an explanation for the apparent inconsistency between lack of managerial skill in the literature and growth in the mutual fund industry. In addition to aligning manager objectives with the benchmark more appropriately, the use of style-consistent benchmarks is also more appropriate for the 'skill versus luck' debate because a style-consistent factor model should provide a better explanation of style-based fund returns (compared to standard factor models) because the factor (benchmark) will be more closely aligned with each fund's objectives and risk return parameters[1]. Chan et al (2009) note that

[1] We are grateful to an anonymous referee for making this point.

the better a benchmark tracks a manager's active portfolio the more confidence there can be that any excess performance is due to skill and not to luck.

In this paper we address these issues by examining the performance of U.S. equity mutual funds, over the period 1990 to 2011, both at the aggregate level and by investment style to determine whether these managers have positive skill, or not. However, we analyse this performance using both standard single, and multi-factor models and by using style-consistent benchmarks produced by Russell Indexes[2]. We then extend the 'luck versus skill debate', by using both multi-factor models and style-appropriate benchmarks. To do this we implement Fama and French's (2010) simulation methodology to determine how much of any fund manager's performance is due to luck (good or bad) and how much is due to skill. Our results indicate that within sub-samples of funds, based on the industry conventions of investment style groups, investment skill exists, as does the antithesis of investment skill. Further, we find that some segments of the market perform much better than others when measured against their respective benchmarks. These segments tend to be in the small-cap and mid-cap sectors where it might be reasonable to expect specialised management skills to be rewarded. Our results also show that the use of standard multi-factor models underestimates managerial ability and overstates the proportion of funds whose abnormal performance can be attributed to chance rather than to skill, when compared against the use of style-consistent, practitioner benchmarks. We also find that standard information criteria measures (such as the Akaike (1973, 1981), and Schwarz (1978) information criteria) for the style-consistent models were mostly lower or comparable to similar measures for the multi-factor models. As a result, we also conclude that style-consistent benchmarks, being better

[2] As Haughton and Pritamani (2005) note, Russell Indexes aim to create indices that maximise the ease with which passive funds and exchange traded funds (ETFs) can be created using their indexes. Huij and Verbeek (2009) highlight the problems of factor models in this respect.

aligned with investment manager mandates, offer a more parsimonious way of accounting for the risks in style and size tilts. Finally, although we find that the addition of the multi-factor risk variables can enhance the explanatory power of performance evaluation models based only an appropriate Russell Style index marginally, the economic meaning on the related signs of the coefficients are often difficult to rationalise.

The paper therefore makes three key contributions to the literature. First, we show that as well as producing results that can be difficult to assign economic meaning, the use of standard multi-factor models underestimate managerial ability and overstate the proportion of funds whose performance can be attributed to chance[3]. In other words, we find that more managers have skill than has been documented elsewhere in the literature. Second, we evaluate different levels of skill across the main style groups typically used in the U.S. mutual fund industry. Finally, we find that the use of style-consistent and investable passive benchmarks provides a parsimonious approach accounting for the size and style tilts of the funds. These findings should have significant implications for both institutional and retail investors when they consider allocating funds to specific sectors of the market and in the decision to select an active or a passive manager. The rest of the paper is organised as follows: in Section 2 we describe our data and methodology; in Section 3 we present our empirical results; while Section 4 concludes the paper.

2. Data and Methodology

In this paper, following industry practice, we examine mutual fund performance among different investment style groups and consider whether the observed performance goes

[3] Chan (2009) concludes that the factor mimicking models "are frequently associated with implausible levels of over- or under-performance".

beyond what can be attributed to ‘luck or chance’. Using monthly returns of over 2,300 US equity mutual funds over the period from 1990 to 2011, we assess fund performance using a range of performance benchmarks suggested in the literature and a simulation technique that follows Fama and French (2010). The range of performance benchmarks includes the use of appropriate Russell Indices that are consistent with the size and style orientation of the respective style groups.

2.1 Data

Our sample of mutual funds is from the Morningstar database and consists of 2,384 surviving and non-surviving U.S. equity funds over the period from January 1990 to December 2011. Our sample selection began with all actively managed funds from the diversified U.S. equity funds sector over the sample period. For the purposes of the regression exercises we only included those funds that should have monthly returns data spanning at least 36 months. In some cases funds may have more than one share class, but these different share classes are not different funds; instead they are based on the same underlying portfolio. These different classes of fund are available to different types of investors; including all of the share classes of all funds, and would not tell us anything additional about manager skill. We deal with this issue by screening for the oldest fund share classes which, as well as having the longest time-series of return (by definition), are generally the largest fund classes and account for the largest proportion of total fund assets[4]. The number of funds in the sample varies from year to year, ranging between 421 funds in 1990 to 1,992 funds at the end of 2011. For each fund, we obtain monthly data on fund returns (both gross and net of fees), assets under

[4] Different authors have addressed this issue in different ways. For example, Vardharaj and Fabozzi (2007) screen out such duplication in their work, retaining only the ‘most popular’ class A shares. In their work Fama and French (2010), combine the different share classes for each fund and create a single fund.

management (AUM) and the fund's equity style as assigned by Morningstar. The Morningstar-assigned equity style box is used to separate funds into different style groups.[5]

The choice of benchmark is an important issue. When examining the performance of fund managers we believe that it is important to use a benchmark that is closely aligned to the manager's active portfolio and mandate. In this paper we use the style-based Russell indices as benchmarks. The Russell indices are the most commonly used industry benchmarks accounting for an estimated 72% market share in terms of U.S. domiciled institutional assets. while the S&P 500 is the most commonly used individual benchmark, and remains the dominant benchmark for the Large-cap Blend segment of the market, it has been overtaken by the Russell Indices for the larger Large-cap segment (according to the Russell Annual Benchmark Survey (2013)). The Large-cap blend segment of the market also contains the vast majority of index funds, which are excluded from our study because we focus only on active managers, who account for 83% of assets under management in U.S. domestic equities. The Russell Style Indices also have a longer history than other style indices, and are commonly used in academic research (see for Chan et al. (2009 and Cremers et al (2012)). We obtained the Russell Total Return Indices used in the style-consistent model from Russell Investments, and the risk factors for the market, size, book-to-market and momentum used in the Fama and French (1993) and the Carhart (1997) models from the Ken French Data Library.[6]

An overview of our data is found in Table 1 which shows the minimum, maximum, average and the standard deviation of the temporal number of funds and fund-months in each style

[5] Morningstar style groups are formed on the basis of size and a combination of the growth and valuation factors which are used to assess a fund's growth/value orientation.
[6] K. French Data library http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Russell Investments http://www.russell.com/indexes/data/US_Equity/Russell_US_equity_indexes.asp.

group over the sample period. There are a total of 355,574 fund-month observations from 2,384 funds over the period from January 1990 through December 2011. The different style groups contain the following number of funds: Large-Cap Growth 389; Large-Cap Blend 488; Large-Cap Value 352; Mid-Cap Growth 253; Mid-Cap Blend 166; Mid-Cap Value 155; Small-Cap Growth 244; Small-Cap Blend 190; and Small-Cap Value 147. All style groups are therefore well represented in the sample. The last column in Table 1 shows the relevant style-consistent Russell Indices used in subsequent analysis.

2.2 Performance Measures

Following the literature, we estimate fund performance using both single-factor and multi-factor asset pricing models. Our empirical framework can therefore be expressed as:

$$\boldsymbol{R_{P,t} - R_{F,t} = \alpha_P + \sum_{k=1}^{K} \beta_{P,k} F_{k,t} + \varepsilon_{P,t}} \tag{1}$$

where $R_{P,t}$ is the fund return, $R_{F,t}$ is the return on the risk free rate, α_P is the fund alpha, $F_{k,t}$ is the k^{th} benchmark or risk factor and $\varepsilon_{P,t}$ is the residual error term.

We estimate alphas from four different perspectives. First, we estimate alphas from the traditional single-factor capital asset pricing model (CAPM) (Jensen (1968)) in which the risk factor is the excess return on the market; second, the Fama and French (1993) three-factor model in which the risk factors are the excess return on the market, plus the Fama and French size and book-to-market factors; third, the Carhart (1997) four factor model where we add to the Fama and French three factors Carhart’s momentum risk factor. The fourth version of the model is a single-factor CAPM-style model in which the risk factor is a style-consistent benchmark return. This version accounts for the fact that fund managers may be constrained in their stock selection by their investment style and objective. The specific Russell Index used for each style group is shown in the last column of Table 1.

In addition to running our models for our full sample and style sub-samples we also run our models for equally-weighted and value-weighted portfolios of all funds and style sub-groups. In calculating the monthly value-weighted portfolio returns, funds in the portfolio are weighted by the assets under management at the previous month-end.

2.3 *Bootstrap Simulations*

Bootstrap experiments have been used extensively in the literature to examine whether superior fund performance is due to managerial skill or pure luck. See for example Kosowski et al, (2006), Fama and French (2010) and Busse et al, (2010). Following Fama and French (2010), we use simulations of individual fund returns to assess whether the observed performance is due to the skill of fund managers or to luck by comparing the distribution of the benchmark-adjusted, actual t(alphas) to the distribution of t(alphas) from equivalent zero-alpha returns. We use the Fama and French (2010) simulation technique because of the concerns raised in their paper about other simulation approaches.

More specifically, we estimate each fund's alpha using a given benchmark model and then subtract the alpha estimate obtained from the fund's monthly returns. For each fund this therefore produces a 'net of alpha' return series, relative to the given benchmark model, that has the same statistical properties as the actual fund returns.- We then randomly select, with replacement, 264 monthly returns for each of the funds. Re-sampling is done in such a way that the 264 months drawn is the same for each fund and the other benchmark returns. This ensures that the cross-sectional correlation between fund returns is preserved in the re-sampled series. Ignoring the cross-sectional dependencies in the fund returns can lead to incorrect inferences. Using the re-sampled series, for each fund we estimate the alpha and t(alpha) for the different benchmarks. Thus for each run the same batch of re-sampled series of zero-alpha returns is used to estimate the alpha and t(alpha) for the benchmark models. The re-sampling process is repeated 10,000 times.

We then examine the distribution of the t(alpha)s from the simulated series and compare that with the distribution of the actual t(alphas) of the individual funds. First, we compare the percentiles of the cross-section of t(alpha) estimates from actual fund returns against the average values of the percentiles from the simulated returns. We then examine the likelihood of the observed percentile t(alphas) of the actual fund returns being realized even when the t(alphas) are known to be zero. This is used to assess whether the observed performance is due to skill or luck.

We focus on the distribution of the t(alphas) instead of the alpha estimates for the same reasons outlined in Kosowski et al (2006) and Fama and French (2010). This approach controls for differences in the precision of the alpha estimates due to differences in the number of observations and the residual variances across the different funds. It should be noted that alpha estimates with low precision could be spurious, albeit economically significant. The t(alphas) can therefore be considered as precision-adjusted alpha estimates.

3. Results

In this section of the paper we present the results of our analysis based on the four benchmark models: the CAPM, Style-Appropriate Russell Indices, the Fama and French three factor model and the four factor model due to Carhart, in several different contexts.

3.1 Aggregate Fund Performance

We begin our empirical analysis by examining the overall, or collective performance of the funds. We do this using equally-weighted (EW) and value-weighted (VW) portfolios of all the funds in our sample. In other words we form two 'portfolios' of mutual funds: the first where each fund has an equal weighting and the second where each fund is weighted according to its NAV. We then use these two aggregated versions of the US equity mutual fund universe as dependent variables in OLS regressions, where the independent variables are the appropriate factors of the four performance models. We estimate each model in gross and net of fee versions. Each regression allows us to say something about the alpha generated by the US equity mutual fund industry over time. These regression results are reported in Table 2.

The equally-weighted CAPM alpha for all funds is 0.13% (0.03%) per month for gross (net) returns with a *t*-statistic of 2.06 (0.47)[7]. The value-weighted portfolio of all funds has a lower alpha of 0.08% (0.01%) per month for the gross (net) returns. These results indicate that any manager skill is concentrated in smaller funds. If we measure performance using the Fama and French three-factor or the Carhart four-factor models we find that there is little value-added in these aggregate portfolios. Only the EW, Gross of fee portfolio based on the three factor model has significant alpha 0.10% (t-statistic = 1.97). But at the net level both the EW portfolio and VW portfolio of funds are negative, though insignificantly so. Thus in summary, funds do generate positive alphas when benchmarked against the one factor model, but only just enough to cover fees because the economic and statistical significance of the alpha disappears when using net returns. These results are largely consistent with the findings

[7] Note that reported alphas in this paper are on percentage per month basis unless otherwise stated in the text and where relevant the t-statistic on the market coefficient tests the null hypothesis that the coefficient equals 1.

reported in the literature (see for example Fama and French (2010)) but we document these results to provide a basis for comparison with subsequent analysis across style groups.

On the basis of equilibrium accounting as espoused by Fama and French (2010), the aggregate results reported in Table 2 could mask any true, value-added or positive alpha performance because positive alpha performance in some style groups may be offset by other fund groups with negative alpha. Kosowski et al (2006) report that while on average funds did not beat their benchmarks there were large sub-groups with strong positive performance, most notably the top 20% of 'Growth' and 'Aggressive Growth Funds'. They also reported the underperformance of funds with the 'Growth and Income' objective, which is typically favoured by value funds. We therefore examine, at the aggregate level, whether there are cross-sectional differences in performance based on investment style. To examine this issue, we construct EW and VW portfolio returns for each style group and estimate their alphas using the four variants of equation 1 above. The results are reported in Table 3, Panels A to D.

In Table 3 Panel A, we present the results based on the CAPM model for nine equity style groups[8]. The gross, equally-weighted alphas are positive for all three Large-cap styles, although none are found to be statistically significant. The net alpha equivalents are all negative, and in the case of the Large-cap Blend category, the alpha is estimated to be -0.07% per month and is statistically significant. The equivalent value-weighted alphas are qualitatively very similar: the gross alphas are positive and small, but still not statistically significant, while the net alphas are all negative although none are found to be statistically significant. The negative net alphas for each category suggest that manager skill is not

[8] The results in the last row in Panels A, B and C of Table 3, representing the results for 'All funds', can also be found in Table 2 and are presented again for ease of reference.

sufficient to cover the fees charged for that skill. The gross and net alphas for the Mid-cap styles are all positive, for both the equally-weighted and value-weighted versions of the results. Relatively speaking, there appears to be more manager skill evident in these styles than in the equivalent Large-cap styles. Furthermore the net alphas are economically significant too. For example, the net alphas for the Mid-cap Value style are 0.15% and 0.12% per month for the equally and value-weighted calculations respectively. However, the alphas are only found to be statistically significant for the equally-weighted, gross Mid-cap Blend and Value styles, though not for the net equivalents, and for the gross value-weighted results for Mid-cap blend style. Finally, the estimated net and gross alphas for the Small-cap styles are all found to be positive for both the equally-weighted and value-weighted calculations. Indeed, the results are qualitatively similar to those for the Mid-cap styles. The net alphas are economically significant, particularly for the Small-cap Value style where we find the equally and value-weighted net alphas to be 0.21% and 0.24% per month respectively. However, once again, we do not find any of the Small-cap net alphas to be statistically different from zero, although the equally-weighted gross alpha for Small-cap Value is found to be statistically significant. Overall, our results indicate scant evidence of significant, positive skill amongst US equity managers when using the traditional CAPM model; any skill that is evident seems to be concentrated in Mid- and Small-cap funds and/or in smaller funds.

Panel B of Table 3 present results based upon the three factor model. The equally-weighted, gross alphas are all positive when we use the three-factor model, although only significantly so for the Mid-cap Blend, Mid-cap Value and Small-cap Value styles. The equivalent net alphas are of course smaller, are negative for all three Large-cap styles, but positive for the Mid-cap and Small-cap styles. However, they are never statistically different from zero. The equivalent value-weighted results are qualitatively similar. As might be expected, the Small-cap funds have a significantly positive exposure to the size factor in Panels B and C, while the Large-Cap funds exhibit a negative exposure to the same factor. For example, in Table 3B the 'Size' coefficient for Small-cap Value funds (equally-weighted) is 0.57 and highly significant while the equivalent size coefficient for Large-cap Blend is -0.07 and again highly significant. Similarly, Growth funds (regardless of size orientation) exhibit a negative exposure to the book-to-market (B/M) factor with Value funds showing a positive exposure. For example Large-cap Growth funds (equally-weighted) have a 'book-to-market' coefficient of -0.20 which is significant, while the same coefficient for Large-cap Value funds is a highly significant 0.25%.

Panel C of Table 3 presents analogous results but using Carhart's four factor model. Again, all of the equally-weighted gross alphas are positive although unlike with the three factor model, none are statistically different from zero. However, the related net alphas are now almost uniformly negative. We find that only the Mid-cap Growth style produces a positive net alpha of 0.02% per month, which is economically small and insignificantly different from zero anyway. The value-weighted net alphas are all found to be negative. With regard to Carhart's differentiating, fourth factor, momentum, the table shows that the coefficients related to this factor are positive in all cases, the Large-cap Growth sector is the exception here, where we find it to be negative and marginally significant for the equally-weighted formulation, but negative and insignificantly different from zero in the value-weighted formulation. The momentum factor is positive, statistically significant, and largest for the three Small-cap styles. For example, for Small-cap Value style funds the equally-weighted momentum coefficient is 0.17 and highly significant.

In Panel D of Table 3 we present estimates based upon the Russell style-consistent benchmarks. First, the evidence of the existence of positive gross alpha (manager skill) is far more compelling when we use style-consistent benchmarks. Second, all of the equally-weighted and value-weighted gross alphas are positive. With regard to the equally-weighted results, of the Large-cap styles only Growth is found to produce a significant gross alpha, but there is evidence of an economically and statistically significant gross Mid-cap Growth and Mid-cap Blend alphas, and even stronger evidence for all three Small-cap styles. The signs of the equally-weighted net alphas relating to Large-cap and Mid-cap Blend and Value styles are negative (but only significantly so for Large-cap Blend). More interestingly, the net alphas for the Small-cap Growth and Blend styles are not only statistically significant, but also economically significant too. The net alphas for Small-cap Growth and Small-cap Blend are 0.28% and 0.17% per month respectively. Again, the value-weighted results are qualitatively similar to the equally-weighted results.

3.2 Aggregate fund performance results summary

The results in Panel D of Table 3 indicate that some managers do exhibit 'skill' when fund performance is measured against a benchmark which is broadly consistent with the investment style and objectives of the fund. Cremers et al (2012) investigate the systematic biases in the three and four factor models which may well play a part in benchmark-varying results, notably the equal weighting of small and large stocks which do not reflect market capitalisation and therefore investment opportunity. However, it is important to note that while Cremers et al (2012) measure the degree of 'active management' inherent in managers' decisions, we investigate where managerial skill lies within the style spectrum. The studies are similar in the sense that both require an appropriate benchmark as a starting point. However, our results support the view that an appropriate benchmark should be style-based since it is these benchmarks that are used to measure and monitor manager performance and, consequently, used to determine manager remuneration.

Our results also show that the choice of benchmark has an impact on the value premium that investors can 'earn' ("value minus growth") and the size premium ("small minus large"). In Panel 3A we find a positive value premium, that is, the alpha of the value portfolio is larger than the alpha of the equivalent growth portfolio. For example, the gross alpha of the Large-Cap Value portfolio is 0.08% while the gross alpha of the Large-cap Growth portfolio is 0.03%. Similarly, Panel 3A shows that the size premium is also positive. For example, the gross alpha of the Small-cap Value portfolio is 0.32%, while the gross alpha of the Large-cap Value portfolio is 0.08%. An interesting result arises when we consider these premia based on the style benchmarks in Panel D of Table 3. First, the size premium remains positive. For example, the gross Small-cap Value alpha is 0.19%, while the gross Large-cap Value alpha is estimated to be 0.04%. However, the value premium becomes negative. For example, the gross Small-cap Value alpha is 0.19% compared with the Small-cap Growth gross alpha of 0.40%; the Mid-cap Value gross alpha is 0.09% compared with the gross Mid-cap Growth alpha of 0.19%; and finally, the gross Large-cap Value alpha is 0.04% compared to the gross Large-cap Growth alpha of 0.12%.

We also find that when we use the three factor model we generate a positive value premium, but the same process using the four factor model produces a negative value premium.[9] When we use the three factor model the size premium is estimated to be negative. Our results show

[9] These observations are based on Table 3 Aggregate Fund Performance. It should be noted that for the CAPM model only the small value style is statistically significant at the gross level whereas for the style appropriate Russell model mid cap growth is statistically significant at the gross level and small cap growth at both the gross and the net level. Like the CAPM model the three factor model is only significant for small value at the gross level whilst no statistically significant performance is recorded at the aggregate level by the four factor model, which incorporates stock price momentum.

that for the three factor model and the style-consistent benchmark model there is some evidence of superior performance amongst some style groups. In aggregate however, where superior performance is observed this is, mostly, just enough to cover fees and expenses.

Overall, the CAPM produces the lowest average R-squared values, particularly with regard to the three Small-cap styles of Growth, Blend and Value which were estimated respectively as 77%, 80% and 75%. Perhaps unsurprisingly the average R-squared values generated for the Small-cap styles using the style-consistent benchmarks were the highest of all four models at 97%, 97% and 96% for the Growth, Blend and Value styles. The equivalent values produced by the four factor model (the next highest) were 96%, 96% and 94% respectively. These results show clearly that the style-consistent benchmarks do as good a job as the multifactor models in explaining fund returns. We can therefore reasonably argue that the respective Russell benchmarks used for each of the style groups adequately captures the risks embedded in the fund characteristics such as size and growth-value orientation. This argument is further supported by the Schwarz Information Criterion (SIC) values reported in Table 4. Table 4 shows that the SICs of the multifactor models are consistently lower than that of the CAPM.[10] The SIC values of the style-consistent Russell Index model, are not only consistently lower than the CAPM but are also lower than the comparative values of the multifactor models in at least six cases out of nine (six in the case of the value-weighed style portfolios and seven in the case of the equally-weighted style portfolios).[11] Thus, the use of the style-appropriate Russell indices may provide a more parsimonious way of accounting for the risks in these style tilts as well as being more in line with industry practice which generally compares fund returns against relevant (ie. industry constructed), passive index benchmarks. Our findings

[10] The SIC is one of the standard approaches to model selection. When comparing the SIC values for different models, the lower the value the better.
[11] Although not reported, the rankings do not change when we use the AIC or the log-likelihood.

should also be of particularly relevance in markets where the appropriate multi-factors are not readily available.

3.3 Combining Russell Style benchmarks with familiar risk factors

Our results in Section 3.1 and 3.2 show how it is possible to use a parsimonious model to evaluate the performance of mutual fund managers, when the risk benchmark is more appropriately aligned with the benchmarks' of the fund managers. In Table 5A we augment the performance evaluation model with the familiar risk factors of Fama and French to see what role these factors have in the presence of the fund-specific Russell Style benchmarks[12]. Essentially we examine the marginal value of the risk factors in the presence of the Russell Style benchmarks. However, given the very high correlation with the "Market" risk factor, to avoid problems of multi-collinearity, we only add the familiar Fama and French size (SMB) and value factors (B/M), along with Carhart's Momentum factor (MOM) to the Russell Style benchmarks.

Panel A in the table presents the results of average OLS coefficients from the re-estimation of expression (1) for net of fee returns[13]. For each style group the table presents average coefficients and related test statistics for estimates of expression (1) where the independent variables are either the relevant Russell Style benchmark, or the relevant Russell Style benchmark plus SMB, B/M and MOM. The second column in Panel A of Table 5A shows that the average alpha for the Large Cap Growth funds is 0.013% per month, the associated t value of 0.83, is calculated from the estimated alphas. The average coefficient on the Russell Large Cap Growth style index is 0.979, with an associated t-value of 91.67. The third column in Panel A presents comparable results for this style group, but with the inclusion of

[12] This additional line of enquiry was suggested by an anonymous referee.
[13] In the interests of completeness Table 5B presents similar results for gross returns.

the SMB, B/M and MOM risk factors. The inclusion of the three risk factors increases the average R^2 from 87.1% to 91.3%, and reduces the average alpha to -0.104%, with an associated t value of (10.01). The average beta on the Russell benchmark increases very marginally to 0.995. The coefficients on all three risk factors representing, the size, value and momentum effect are positive and highly significant.

The results in Panel A of Table 5A for the Large Cap Blend and Mid Cap Growth funds are qualitatively similar to those for the Large Cap Growth funds. However, the results for the remaining six categories are different. For these style groups one or more of the average coefficients on the risk factors are negative. For example, the results for the Large Cap Value style group show that the average coefficients on both the Value (B/M) and the Momentum factors (MOM) are both negative, at -0.114 and -0.061 respectively, and are both highly significant with t values of 10.07 and 9.46 respectively. Furthermore the intuition behind the factors is not supported by the coefficient values. For example, the size factor (SMB) is positive for all categories, including all the Large Cap styles, but is negative for the Small Cap Growth style. Similarly the Value risk factor (B/M), is positive for seven of the categories, including all three growth styles, but is negative for the Large and Mid Cap Value styles. A simple conclusion that can be drawn from the varying and 'incorrect' signs of the coefficients on the risk factors when added to the Russell indices is that style tilts have already been accommodated by the index and that the factors are essentially surplus to requirement.

Panel B of Table 5A presents results that are equivalent to those in Panel A, but where the independent regressors – the return on the Russell indices, and on the risk factors – have been rescaled by dividing each independent variable by its own standard deviation. This rescaling does not affect the estimates of alpha, or the t values on the average coefficients, or R^2 of the regressions, but does change the scale of the average coefficients on the returns on the Russell Style benchmark and on the additional risk factors. For example, the average coefficient on the Russell Style benchmark for the Large Cap Growth style group is 0.979 in Panel A but 4.94 in Panel B. Although this rescaling does not change the significance of the estimated alphas or the independent regressors it does help to put into perspective the marginal explanatory value of each of the independent regressors.

The coefficients presented in Panel B allow for a clear comparison of the impact on fund return due to a one standard deviation increase in the value of a regressor in the model. For example, for the Large Cap Growth style the average coefficient on the Russell Index Benchmark (RS) is 4.94 in column 2, and 5.02 in column 3. This suggests that a one standard deviation shock to the returns on the benchmark leads to an increase or decrease in fund returns of around ±5% (depending upon the direction of the shock). In column 3, the average coefficients of the SMB, B/M and MOM risk factors are 0.411, 0.289, and 0.53 respectively. Which in turn means that a one standard deviation shock in these risk factors leads to a ±0.411%, ±0.289% and ±0.53% in fund returns respectively. Although each of the coefficients on the additional market risk factors is statistically significant, the impact of the benchmark returns on fund returns on average is a minimum of around 10 times larger in the Large Cap Growth style category. In some style categories, for example for the Large Cap Blend group, the impact of a one standard deviation shock in the style benchmark has an impact on fund returns that is more than 25 times the impact of the same scale shock to any of the risk factors.

These results demonstrate that the role that the well-known risk factors play in explaining returns is marginal in the presence of the style benchmarks. Furthermore, as stated earlier we do not believe that these factors represent appropriate tools for evaluating the performance of fund managers given that they are not benchmarked against these factors by their managers or by their clients, and because these factors, in contrast to the Russell indices, are neither investible benchmarks and are unavailable in most mutual fund markets.

3.3 Luck versus skill in performance

Although the results presented in Section 3 suggest that fund managers can outperform some benchmarks in some style groups (particularly within the Small-cap style), they do not necessarily provide evidence of 'managerial skill'. As argued by Fama and French (2010), among others, the outperformance could be due to good fortune. After all, bad managers can be lucky while good managers can be unlucky. We now explore this issue.

We apply the bootstrap simulation approach, suggested by Fama and French (2010) to examine whether there is evidence that superior manager skill is style group-specific. We do this in two steps. First, we examine whether funds on average exhibit superior skill; in this case we aggregate all the funds and ignore any differential performance across style groups. Second, we conduct the same exercise but on all funds within the nine Morningstar style groups which correspond to commonly used market segments. This enables us to make inferences about skill across the different style groups. In the interests of parsimony our discussion of the results focuses on the results generated by the three-factor model and on the style-consistent, single-factor benchmark models[14].

[14] Those results based upon the CAPM and the four factor model which are equivalent to those presented in Tables 5, 6 and 7 are presented in appendices 1, 2 and 3.

Following Fama and French's methodology, in Tables 6 to 8 we present actual fund t(alpha)s and simulated t(alpha)s at various percentile breakpoints in order to consider whether any extreme negative or positive performance is due to skill or luck. In the interests of parsimony all results are based on net of fee fund returns. The simulated t(alphas)s have been generated using the procedure described in Section 2 of this paper. More precisely we present estimates of t(alpha) at selected percentiles of the cumulative distribution function (CDF) for the actual net returns (Actual), the average of the 10,000 simulations (Sim), and in the column headed '%<Actual', the percentage of the simulated runs that produce lower values of t(alphas) at the selected percentiles than the observed percentile value for the actual fund returns. Statistically significant results are highlighted in bold print in the tables. Finally, in Figures 1 to 3 we present pictures of the full, related CDFs for each distribution of t(alphas) presented in Tables 6, 7 and 8.

In Table 6 Panel A, we present the aggregate results for the three factor model for gross and net of fee returns. The related cumulative distribution functions (CDF) for the actual and simulated gross and net t(alphas) are shown in full in Panel A of Figure 1. The net of fee results indicate evidence that the bottom five per cent of managers have demonstrated negative skill. At the other end of the performance spectrum, only the top two per cent of managers generate positive alpha that is due to skill (net of fees); positive alpha produced elsewhere is not distinguishable from luck. These results are consistent with those of Fama and French (2010). The results are also consistent with those of Busse et al (2010) who find that the distribution of actual t(alpha) and simulated t(alpha) for the four factor model were indistinguishable from their simulated counterparts. However, Kosowski et al (2006) reported that while on average funds did not beat their benchmarks there were large sub-groups with strong positive performance, most notably the top 20% of 'Growth' and

‘Aggressive Growth Funds’. They also reported underperformance by funds with the ‘Growth and Income’ objective, which is typically favoured by value funds.[15]

In Panel B of Table 6 we present the aggregated actual and simulated t(alphas) based upon the style-consistent benchmarks, and present the corresponding CDFs in full in Panel B in Figure 1. Again we focus here on the net of fee results. Using the style-consistent benchmarks we find evidence again that the bottom five per cent of managers have demonstrated value reducing skill. However, at the other end of the performance spectrum we find that five per cent of managers to have added value to their funds’ performance, compared with the just two per cent of managers when we use the three factor model.

To summarise, the results based on the style-consistent benchmarks indicate stronger evidence of skill at the top end, while at the lower end the net of fee performance is still due to value reducing managerial skill. At the aggregate level then our results are similar to those of Fama and French (2010) and support their equilibrium accounting view of mutual fund performance. Given our emphasis on the role of style benchmarks and our belief in their potency for investment mandates and performance measurement, what can we say about luck versus skill within different style groupings?

3.4 Luck versus skill in style group performance–net of fees performance

Table 7 shows the net of fee, bootstrapped results based on the three factor Fama-French model applied to the different Morningstar styles. Figure 2 shows the related CDFs. For all Large-cap fund styles we find the same pattern. In the left hand tail (bottom 10%) there is evidence of ‘wealth reducing skill’ since the simulated alphas are lower than their actual

[15] Kosowski et al (2006) consider four prospectus investment objectives Aggressive Growth, Growth, Growth and Income and Balanced or Income Funds. Such broad categorisation is too general to be useful in practice and Balanced or Income Funds include significant proportions of securities e.g. fixed income which are not aligned with the benchmarks applied.

equivalents. At the top end of the performance spectrum, where positive t(α) is observed, it is almost certainly due to luck, since simulated t(alpha)s are always higher than equivalent actual alphas.

The results for the Mid-cap funds are less straightforward. For the Mid-cap Growth funds we find evidence of a lack of skill since the actual t(α)s are more negative than the simulated equivalents. However, the top two per cent of managers in this category demonstrate statistically significant evidence of positive skill. For Mid-cap Blend funds we find evidence of significant negative skill at the lower end of the performance range (bottom 5 per cent), but evidence of significant, positive value-added skill for the top 5 per cent of performers. The results of the Mid-cap Value funds is very similar to that of the equivalent Blend funds, where the top five per cent of managers demonstrate skill, and the bottom 4 per cent demonstrating significant negative skill.

The Small-cap results in Table 7 also produce some interesting results. For Small-cap Growth funds we find evidence of negative, or value reducing skill in the lower tail, but evidence that luck plays a large role in any apparent positive skill found in the upper tail. Only the top 1% of Small-cap Growth managers seem to have skill. For Small-cap Blend funds we find similar results, that is, evidence of negative skill at the bottom end of the performance spectrum (bottom 5 per cent), and evidence of a limited amount of positive skill among the top 3 per cent of managers. For Small-cap Value funds we find less evidence of negative skill at the bottom end of the performance range with only the bottom two per cent of managers demonstrating negative skill. We find that evidence of positive skill can only be found for the top 4% of performers of Small-cap Value funds.

In Table 8 we present results based on the style-consistent Russell indices (e.g. Russell 1000 Growth for large-cap growth) as benchmarks; Figure 3 presents the related CDFs. When we consider performance against the Russell Indices (net of fees) it is immediately apparent that there are considerable differences with regard to the degree of skill evident in fund performance. Broadly speaking as we move from funds invested in Large-cap stocks through to Mid-cap and small-cap stocks we see progressively more evidence of skill. Our results are consistent with the comments of Busse et al (2010) who note that while it may be difficult to observe skill when controlling for various factor models there is considerable heterogeneity in performance. This is clearly illustrated when we consider the performance of mutual funds on a style basis. Our results suggest strong evidence of skill at the net level for small-cap funds as illustrated in Figure 2 Cumulative Distribution Functions by Style (based on the results outlined in Table 8). The bottom set of plots shows a clear divergence between actual and predicted performance for Small-cap funds.

For Large-cap Growth funds we find evidence of the antithesis of skill within the bottom 5% of performers and evidence that luck plays a big role in those managers in the 99th percentile that produce positive alpha. For the Large-cap Blend funds at the lower levels, the bottom 10%, we see that fund managers show significant evidence of value destroying skill, while at the very upper end of the scale (top 1%) where positive performance is found, we find that this positive performance is probably due more to luck than to skill. For Large-cap Value funds we again find significant evidence of the antithesis of skill in the lower tail (bottom 5%) while at the upper tail any apparent evidence of skill in the top two per cent of funds is once again probably due to luck rather than to skill.

Our results for the Mid-cap Growth funds show that in the upper tail (top 2%) the positive performance appears to be determined by skill, whereas value destroying skill is the source of underperformance in the lower tail. We find that the bottom ten per cent of Mid-cap Blend funds demonstrate negative skill, while the top 3% demonstrate evidence of manager skill. We find similar, though less extreme evidence when we consider the performance of Mid-cap Value funds. The bottom five per cent of Mid-cap Value managers demonstrate evidence of negative skill, while the top two per cent of these managers demonstrate evidence of skill.

Finally, the style-consistent results for the Small-cap funds presented in Table 8 indicate less evidence of negative skill; only the bottom 3 per cent of Small-cap Blend managers demonstrate negative skill, and there is no evidence of negative skill amongst Small-cap Growth or Value managers, poor performance here seems to be due more to bad luck. Second, we find clear evidence of positive skill for the top ten per cent of all Small-cap styles. This evidence of greater skill among this group of managers is perhaps demonstrated more clearly in Figure 3 which shows that the actual t(alpha) is considerably higher than for the simulated results in more than 99% of cases. These results are consistent with those of Schultz (2010) who finds evidence of stock picking skill among Small-cap growth managers.

3.5 Results summary

Overall, we see more evidence of positive skill when moving from the three-factor Fama and French model to the use of style-consistent benchmarks[16]. These results lead us to the conclusion that more generic factor models may be mis-specifying the levels of luck or skill which exists within the mutual fund industry. What general points can be extracted from these results? First, we find that when using appropriate style benchmarks there is more

[16] See also the equivalent and largely comparable results for the four factor model presented in Appendix 3.

evidence of manager skill, that is, when compared against their mandates/prospectuses, US equity mutual fund managers seem to possess some skill. Argon and Ferson (2006), Sensoy (2009), Chan et al, (2009), Cremers et al (2012) and Angelidis et al (2013) all argue that a dedicated style-consistent passive benchmark should provide a more accurate, and more appropriate estimate of a manager's value-added skill. In addition, we believe that it is more appropriate to evaluate fund manager performance against benchmarks that better reflect managers' goals. When we compare our Russell Style benchmark results with those of Fama and French (2010) we find that our results are not as extreme as their CAPM-based results, that is, not as bad at the bottom end or as strong at the top end, presumably because the Russell Style benchmarks are more closely aligned to funds' investment universes and objectives.

Second, we find that there are different levels of skill to be found across the main style groups within the U.S. mutual fund industry. In the Large-cap segment of the market, where information content and analyst coverage is very high it could be argued that more luck is needed to differentiate fund managers from the pack. This, and economies of scale, may in part explain the concentration of index funds in this area.

Third, in the Mid-cap and Small-cap segments of the US mutual fund industry, where we might reasonably expect proprietary, fundamental analysis to yield more benefits, there is statistically significant evidence of skill not only when we consider performance using the single factor, style-consistent indices, but also when we use the three factor model (and the four factor model). Style groups and market segments where stock picking skill have been noted have tended to be Small-cap or growth where keeping ahead in the 'information race' can reward the required diligence. Evidence of manager skill in these areas would also be consistent with the concept of the existence of ‘private information’ as considered by Grossman and Stiglitz (1980), and Cullen et al, (2010) and Schultz’ (2010) explanation of the economic rationale for these segments of outperformance in terms of market efficiency, including liquidity, the costs of gathering information and the level of analytical skill required. This value-added performance among Small-cap managers is achieved despite higher management fees than those charged by Large-cap managers where economies of scale can be reaped (see ICI (2013)). While we find evidence of skill in the Small-cap and Mid-cap sectors we find no such evidence for Large-cap funds where any significant alpha is probably due to luck and not to skill.

The results we have reported show that Mid-cap and Small-cap styles seem to perform better than Large-cap styles against the performance models we have used at the upper end of the performance spectrum, but the performance of Mid-cap and Small-cap blend funds is generally worse at the lower end of the spectrum.

Another point that we would like to emphasise strongly is that the results presented in Tables 6 and 7 suggest that the risks of getting things ‘right or wrong’ are greater for Blend funds than for those that stick to one style. For example, whether using the style-consistent or three factor model the bottom ten per cent of Large Blend managers demonstrate significant negative skill. Perhaps we can infer from this that it is better to be a specialist rather than a generalist? Although, the ability to blend growth and value styles might seem attractive, it may be that timing moves from one style to another is difficult and that the rigour associated with sticking to a clear style discipline might be preferable.

When we introduce the Small Minus Large (SML) and High Minus Low (B/M) factors of the Fama and French (1992) model we find performance differences that probably tell us as much about the factors themselves as about individual styles (see Table 3 Panel C or Cremers et al, (2012)). This is perhaps most evident when we consider the performance of the Small-cap managers. The style-consistent results for the Small-cap Growth and Value funds in Table 8, indicate that performance at the lower end of the performance spectrum might be more the result of bad luck. However, when we look at the same tail and group of Small-cap funds using the three-factor model in Table 7, which has a 'size factor', we find stronger evidence of negative skill. At the top end of the performance spectrum, the results based upon the three factor model for the Small-cap stocks (Table 7) indicate that most of this positive performance is due to luck, not skill. But when we use the style-consistent benchmarks we find strong evidence of manager skill (Table 8). We therefore conclude that the standard, multifactor models not only underestimate managerial ability, but also overstate the proportion of funds whose performance can be attributed to chance.

4. Conclusions

Using both industry and style-consistent benchmarks we have considered the role of skill and luck in the performance of US equity mutual funds and asked the question: does style matter. The academic literature is firmly wedded to the use of multifactor benchmarks whereas in practice fund managers are generally mandated to benchmark the performance of their funds against commercially available style-consistent ones, the most important of which by industry penetration are constructed by Russell Indexes. We discover economically and statistically significant performance differences when we use typical factor models compared with style-consistent benchmarks when applied to different market segments of the U.S. mutual fund industry. We therefore conclude that the selection of an appropriate performance benchmark should be a vital consideration in the assessment of manager investment skill. Our results support the view that looking at mutual funds by market segment or investment style provides the investor or investment sponsor with considerably more information about the existence of value-added skill, or luck, than is revealed at the aggregate level using standard multi-factor models. Although the average fund in our aggregate universe does little more than cover its costs we find considerable variability in results when analysed by style group with small proportions of funds exhibiting both value-added skill and value-destroying behaviour. The U.S. mutual fund market is a differentiated market where managers and their clients consider managers' performance on the basis of style peer groups and passive benchmarks that possess similar risk- return characteristics to the funds' objectives. We find that style differentiation of performance conveys useful and accurate information about the skill or lack of skill of investment managers. Our findings also indicate that the use of style-consistent and investable benchmarks provide a parsimonious way of assessing performance measurement which accounts for the size and style tilts of U.S. equity mutual funds.

Our results indicate that the standard multifactor models which are most often used in academic studies understate the existence of skill and overstate the role of luck in excess returns, or alpha generation. Further, when we disaggregate the equity mutual fund universe there is sufficiently diverse information based on fund styles to warrant careful evaluation and due diligence in the selection of funds within and between styles. The economic implications of these findings are substantial with the worst performing funds recording a negative t(alpha) of around 5%, while the very best funds record a positive t(alpha) of more than 3% against their relevant benchmark index. This information is economically useful, and may in part influence the investor decision as to whether or not they use an active fund manager or a passive manager.

References

Akaike, H.,1974. A new look at the statistical model identification. IEEE Transactions on Automatic AC 19, 716-723.

Akaike, H.,1981. Likelihood of a model and information criteria. Journal of Econometrics 16, 3-14.

Angelidis, T., Giamouridis, D., Tessaromatis, N., 2013. Revisiting mutual fund performance evaluation. Journal of Banking and Finance 37(5), 1759–1776.

Aragon, G.O., Ferson, W.E., 2006. Portfolio performance evaluation. Foundations and Trends in Finance 2 (2), 83-190.

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. Journal of Finance 65(1), 179–216.

Berk, J. B., and I. Tonks, 2009. Return Persistence and Fund Flows in the Worst Performing Mutual Funds. NBER Working Paper 13042.

Busse, J.A., Goyal, A., Wahal, S., 2010. Performance persistence in institutional investment management. Journal of Finance 65(2), 765–790.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 56 , 45-85.

Chan, L. K. C., Chen, H. L., Lakonishok, J., 2002. On mutual fund investment styles. Review of Financial Studies, 15, 1407–1437.

Chan, L., K., Dimmock, S. G., Lakonishok,J., 2009. Benchmarking money manager performance: Issues and evidence. Review of Financial Studies 22(11), 4553-4599.

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

Christopherson, J.A., Carino, D.R. and Ferson W.E., 2009. Portfolio Performance Measurement and Benchmarking. Mc Graw Hill.

Clarke, R., De Silva, H., & Thorley, S., 2002. Portfolio constraints and the fundamental law of active management. Financial Analysts Journal 58(5), 48-66.

Cremers, M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. The Review of Financial Studies 22(9), 3329-3365.

Cremers, M., Petajisto, A., Zitzewtz, E., 2012. Should benchmark indices have alpha? Revisiting performance evaluation. Critical Finance Review 2, 1–48.

Cullen, G., Gasbarro, D., Monroe, G. S., 2010. Mutual fund trades and the value of contradictory private information. Journal of Banking & Finance 34(2), 378-387.

Cuthbertson, K., Nitzsche, D., O'Sullivan, N., 2008. UK Mutual Fund Performance: Skill or luck? Journal of Empirical Finance 15(4), 613-634.

Fama, E. F., French K, R., 1992. The cross-section of expected stock returns. Journal of Finance 47(2), 427-465.

Fama, E. F., French K, R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33(1), 3-56.

Fama, E. F., French K, R., 2010. Luck versus skill in the cross-section of alpha estimates. Journal of Finance (65)5 , 1915–1947.

Fama, E., F., Kenneth R. French K., R., 2012. Size, value, and momentum in international stock returns. Journal of Financial Economics 105(3): 457-472.

Federal Reserve Statistical Release H.8, Selected Assets and Liabilities of Commercial Banks in the United States. August 30, 2013.

Goetzmann, W., and R. Ibbotson, 1994. Do Winners Repeat: Predicting Mutual Fund Performance. The Journal of Portfolio Management 20, 9-18.

Grinblatt, M., and S. Titman. 1992. The Persistence of Mutual Fund Performance. Journal of Finance 47, 1977-1984.

Grinblatt, M., Titman, S., 1993. Performance measurement without benchmarks: An examination of mutual fund returns. Journal of Business 66(1), 47-68.

Grossman, S., J. Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. American Economic Review 70: 393–408.

Haughton, K., and Pritamani, M., 2005. US Equity Style Methodology. Russell Research Commentary, August 2005.

Huij, J., and M., Verbeek, 2009. On the use of multifactor models to evaluate mutual fund Performance', Financial Management 38:1, 75-102.

The Investment Company Institute, 2013. 2013 Investment company fact book: A review of trends and activities in the U.S. investment company industry 53rd ed.

Jensen, M., 1968. The Performance of Mutual Funds in the Period 1945-1964. Journal of Finance 23,(2): 389-416

Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund 'stars' really pick stocks? New evidence from a bootstrap analysis. Journal of Finance (61) 6, 2551-2595.

Kothari, S. P., Warner, J. B., 2001. Evaluating mutual fund performance. Journal of Finance 56(5), 1985-2010.

Menzly, L., Ozbas, O., 2010. Market segmentation and cross-predictability of returns. Journal of Finance 65(4), 1555-1580.

Schultz, P., 2010. Rational cross-sectional differences in market efficiency: Evidence from mutual fund returns. Journal of Financial & Quantitative Analysis, 45(4),847-881.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461-464.

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

Vardharaj, R., Fabozzi, F., 2007, Sector, style, region: Explaining stock allocation performance. Financial Analysts Journal, 63(3), 59-70.

Wermers, R., 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. Journal of Finance, 55(4), 1655-1695.

Table 1: Distribution of Funds by Style Over the Sample Period

This table presents the minimum, maximum, average and standard deviation of the number of funds in each style group over the full sample period. It also shows the total number of fund-months and the Russell indices used as appropriate style-consistent benchmarks for the respective style groups.

Style Group	Minimum	Maximum	Average	Total	Standard Deviation	Fund-months	Style-consistent benchmark
Large-cap Growth	57	335	212.5	389	98.5	56,097	Russell 1000 Growth
Large-cap Blend	98	398	272.7	488	98.9	72,000	Russell 1000
Large-cap Value	77	305	207.7	352	78.3	54,823	Russell 1000 Value
Mid-cap Growth	42	229	150.6	253	65.0	39,761	Russell Mid Cap Growth
Mid-cap Blend	37	144	95.9	166	33.7	25,307	Russell Mid Cap
Mid-cap Value	26	140	80.7	155	40.2	21,306	Russell Mid Cap Value
Small-cap Growth	26	222	140.0	244	66.8	36,968	Russell 2000 Growth
Small-cap Blend	11	173	100.2	190	54.6	26,461	Russell 2000
Small-cap Value	23	135	86.6	147	39.1	22,851	Russell 2000 Value
All Funds	**397**	**2,070**	**1,346.9**	**2,384**	**572.4**	**355,574**	

Table 2: Regression coefficients for equally-weighted and value-weighted portfolios of funds on a gross and net return basis (1990 - 2011)

This table reports the regression coefficient estimates (coef) and corresponding t-statistics (t(coef)) for the CAPM, three-factor, and four-factor versions of regression (1) estimated on net and gross returns on the equal-weighted (EW) and value-weighted (VW) portfolios of funds in our sample. The t(coef) for the variable 'Market' tests whether the coefficient is different from 1. The table also reports the adjusted R-squared (RSQ). The sample consists of 2,384 diversified US equity mutual funds and covers the period from January 1990 to December 2011. *, ** and *** indicate significance at the 90%, 95% and 99% level respectively.

		α_{Gross}	α_{Net}	Market	Size	B/M	Mom	RSQ
EW Returns								
CAPM	coef	0.13**	0.03	1.01				0.96
	t(coef)	(2.06)	(0.47)	(0.58)				
3F	coef	0.10**	-0.01	0.97***	0.21***	0.02		0.98
	t(coef)	(1.97)	(-0.12)	(-2.65)	(10.90)	(1.13)		
4F	coef	0.05	-0.05	0.98*	0.22***	0.06***	0.05**	0.98
	t(coef)	(1.08)	(-1.00)	(-1.67)	(11.83)	(2.67)	(2.28)	
VW Returns								
CAPM	coef	0.08*	0.01	0.99				0.98
	t(coef)	(1.83)	(0.14)	(-1.10)				
3F	coef	0.06	-0.01	0.97***	0.08***	0.01		0.99
	t(coef)	(1.60)	(-0.21)	(-2.69)	(5.95)	(0.93)		
4F	coef	0.03	-0.04	0.98*	0.09***	0.04***	0.04**	0.99
	t(coef)	(0.77)	(-0.97)	(-1.88)	(7.13)	(3.12)	(2.33)	

Table 3A: Aggregate fund performance by style groups

The table presents monthly alphas, factor coefficients and their corresponding t-statistics for the aggregate portfolios of the different fund style groups. The t-statistics are shown in parenthesis. In all panels, the t-statistics on the variable 'Market' or the variable 'Russell' test the hypothesis that the associated coefficient equals 1. Panel A presents results for the single factor CAPM version, Panels B and C present results for Fama and French (1993) three-factor and the Carhart (1997) four factor versions and Panel B presents results for the single factor style-consistent benchmark version. *, ** and *** indicate significance at the 90%, 95% and 99% level respectively.

CAPM Version

	Equally-weighted				Value-weighted			
Fund Group	α_{Gross}	α_{Net}	**Market**	**RSQ**	α_{Gross}	α_{Net}	**Market**	**RSQ**
Large-cap Growth	0.03	-0.07	1.07***	0.95	0.01	-0.07	1.10***	0.92
	(0.32)	(-0.77)	(3.03)		(0.08)	(-0.66)	(3.70)	
Large-cap Blend	0.01	-0.07**	0.94***	0.99	0.03	-0.03	0.94***	0.99
	(0.36)	(-2.30)	(-5.10)		(0.78)	(-0.87)	(-3.74)	
Large-cap Value	0.08	-0.01	0.85***	0.89	0.08	0.01	0.86***	0.89
	(0.81)	(-0.06)	(-4.12)		(0.74)	(0.14)	(-3.89)	
Mid-cap Growth	0.17	0.06	1.17***	0.86	0.14	0.05	1.20***	0.86
	(1.24)	(0.45)	(5.06)		(1.02)	(0.37)	(5.66)	
Mid-cap Blend	0.21**	0.10	1.00	0.93	0.19*	0.10	1.05*	0.93
	(2.15)	(1.03)	(0.08)		(1.93)	(1.06)	(1.79)	
Mid-cap Value	0.25*	0.15	0.87***	0.84	0.20	0.12	0.87**	0.81
	(1.71)	(1.01)	(-2.82)		(1.23)	(0.75)	(-2.20)	
Small-cap Growth	0.21	0.08	1.21***	0.78	0.20	0.10	1.20***	0.77
	(1.18)	(0.47)	(5.97)		(1.12)	(0.57)	(5.46)	
Small-cap Blend	0.26	0.14	1.05	0.81	0.22	0.13	1.04	0.80
	(1.62)	(0.91)	(1.15)		(1.36)	(0.81)	(1.14)	
Small-cap Value	0.32*	0.21	0.92	0.76	0.32*	0.24	0.92	0.75
	(1.71)	(1.13)	(-1.40)		(1.67)	(1.22)	(1.35)	
All Funds	0.13**	0.03	1.01	0.96	0.08*	0.01	0.99	0.98
	(2.06)	(0.47)	(0.58)		(1.83)	(0.14)	(1.10)	

Table 3B: Aggregate fund performance by style groups (see Table 3A for notes)

Fama & French 3-Factor Model

	Equally-weighted						Value-weighted					
Fund Group	α_{Gross}	α_{Net}	**Market**	**Size**	**B/M**	**RSQ**	α_{Gross}	α_{Net}	**Market**	**Size**	**B/M**	**RSQ**
Large-cap Growth	0.05	-0.04	1.06***	0.04	-0.20***	0.97	0.04	-0.04	1.08***	0.08***	-0.25***	0.97
	(0.88)	(-0.70)	(3.55)	(1.50)	(-7.51)		(0.52)	(-0.58)	(4.18)	(2.85)	(-9.26)	
Large-cap Blend	0.01	-0.07***	0.96***	-0.07***	0.05***	0.99	0.03	-0.03	0.95***	-0.05***	0.03*	0.99
	(0.55)	(-2.70)	(-5.36)	(-4.84)	(3.53)		(0.96)	(-0.88)	(-3.51)	(-3.50)	(1.67)	
Large-cap Value	0.06	-0.03	0.88***	-0.10***	0.25***	0.96	0.06	-0.01	0.90***	-0.12***	0.25***	0.96
	(0.99)	(-0.51)	(-6.32)	(-3.40)	(9.09)		(0.96)	(-0.11)	(-5.42)	(-4.87)	(11.03)	
Mid-cap Growth	0.15	0.04	1.08***	0.41***	-0.27***	0.95	0.12	0.03	1.10***	0.42***	-0.28***	0.96
	(1.64)	(0.46)	(3.19)	(12.34)	(-8.74)		(1.36)	(0.36)	(4.18)	(12.49)	(-9.49)	
Mid-cap Blend	0.15**	0.04	0.96**	0.25***	0.11***	0.96	0.14*	0.05	1.006	0.24***	0.10***	0.96
	(2.18)	(0.62)	(-1.97)	(8.50)	(3.99)		(1.80)	(0.67)	(0.27)	(9.43)	(4.24)	
Mid-cap Value	0.18**	0.08	0.86***	0.13***	0.31***	0.93	0.13	0.05	0.86***	0.11***	0.36***	0.91
	(2.01)	(0.86)	(-4.97)	(2.74)	(9.08)		(1.26)	(0.49)	(-3.74)	(2.55)	(10.94)	
Small-cap Growth	0.13	0.01	1.05**	0.76***	-0.24***	0.96	0.12	0.03	1.036	0.77***	-0.25***	0.96
	(1.45)	(0.07)	(2.32)	(22.26)	(-9.84)		(1.40)	(0.29)	(1.51)	(17.13)	(-7.08)	
Small-cap Blend	0.14	0.03	0.93***	0.64***	0.13***	0.95	0.10	0.01	0.92***	0.68***	0.10***	0.95
	(1.64)	(0.30)	(-2.90)	(16.58)	(4.20)		(1.21)	(0.13)	(-3.21)	(20.61)	(3.87)	
Small-cap Value	0.18**	0.07	0.82***	0.57***	0.33***	0.93	0.18*	0.10	0.82***	0.57***	0.34***	0.93
	(2.02)	(0.82)	(-6.07)	(11.14)	(7.77)		(1.90)	(1.00)	(-5.67)	(13.08)	(8.36)	
All Funds	.10**	-0.01	0.97***	0.21***	0.02	0.98	0.06	-0.01	0.97***	0.08***	0.01	0.99
	(1.97)	(-0.12)	(-2.65)	(10.90)	(1.13)		(1.60)	(-0.21)	(-2.69)	(5.95)	(0.93)	

Table 3C: Aggregate fund performance by style groups (see Table 3A for notes)

Carhart 4-Factor Model

	Equally-weighted							Value-weighted						
Fund Group	α_{Gross}	α_{Net}	**Market**	**Size**	**B/M**	**Mom**	**RSQ**	α_{Gross}	α_{Net}	**Market**	**Size**	**B/M**	**Mom**	**RSQ**
Large-cap Growth	0.09	-0.01	1.05***	0.03	-0.23***	-0.04*	0.97	0.07	-0.01	1.07***	0.07***	-0.28***	-0.03	0.97
	(1.49)	(-0.10)	(2.96)	(1.25)	(-7.77)	(-1.74)		(0.96)	(-0.12)	(4.03)	(2.76)	(-9.66)	(-1.24)	
Large-cap Blend	0.02	-0.07**	0.96***	-0.07***	0.04***	0.00	0.99	0.03	-0.04	0.96***	-0.05***	0.04***	0.01	0.99
	(0.52)	(-2.41)	(-4.96)	(-4.80)	(3.57)	(-0.10)		(0.66)	(-0.97)	(-3.34)	(-3.68)	(2.93)	(0.62)	
Large-cap Value	0.01	-0.08	0.90***	-0.09***	0.30***	0.05	0.96	0.00	-0.06	0.91***	-0.11***	0.30***	0.06*	0.97
	(0.19)	(-1.52)	(-7.85)	(-3.16)	(8.94)	(1.59)		(0.05)	(-1.20)	(-6.52)	(-4.83)	(10.09)	(1.80)	
Mid-cap Growth	0.13	0.02	1.09***	0.41***	-0.24***	0.03	0.95	0.09	0.00	1.11***	0.42***	-0.25***	0.04	0.96
	(1.36)	(0.20)	(3.25)	(12.26)	(-6.14)	(0.83)		(0.94)	(-0.03)	(4.27)	(13.08)	(-7.28)	(1.24)	
Mid-cap Blend	0.10	-0.01	0.973	0.26***	0.16***	0.06*	0.96	0.10	0.02	1.016	0.24***	0.14***	0.04	0.96
	(1.41)	(-0.16)	(-1.36)	(8.89)	(4.40)	(1.76)		(1.35)	(0.22)	(0.78)	(9.67)	(3.92)	(1.29)	
Mid-cap Value	0.10	-0.01	0.88***	0.15***	0.39***	0.09**	0.93	0.05	-0.03	0.89***	0.13***	0.44***	0.09*	0.92
	(1.21)	(-0.08)	(-5.34)	(3.12)	(8.55)	(2.11)		(0.51)	(-0.37)	(-3.83)	(2.92)	(8.48)	(1.88)	
Small-cap Growth	0.06	-0.07	1.07***	0.78***	-0.17***	0.08**	0.96	0.04	-0.06	1.06**	0.79***	-0.17***	0.09***	0.96
	(0.64)	(-0.77)	(2.95)	(24.92)	(-4.68)	(2.55)		(0.48)	(-0.67)	(2.24)	(18.53)	(-4.18)	(2.64)	
Small-cap Blend	0.00	-0.11	0.97*	0.68***	0.26***	0.15***	0.96	-0.03	-0.12	0.95**	0.71***	0.22***	0.14***	0.96
	(0.06)	(-1.40)	(-1.83)	(20.92)	(5.89)	(3.16)		(-0.38)	(-1.62)	(-2.25)	(29.53)	(5.26)	(3.10)	
Small-cap Value	0.03	-0.08	0.87***	0.61***	0.48***	0.17***	0.94	0.02	-0.07	0.87***	0.61***	0.50***	0.18***	0.94
	(0.38)	(-1.02)	(-7.68)	(12.85)	(9.89)	(3.49)		(0.24)	(-0.86)	(-6.72)	(16.31)	(10.32)	(3.85)	
All Funds	0.05	-0.05	0.98*	0.22***	0.06***	0.05**	0.98	0.03	-0.04	0.98*	0.09***	0.04***	0.04**	0.99
	(1.08)	(-1.00)	(-1.67)	(11.83)	(2.67)	(2.28)		(0.77)	(-0.97)	(-1.88)	(7.13)	(3.12)	(2.33)	

Table 3D: Aggregate fund performance by style groups (see Table 3A for notes)

Russell Style Benchmark Models

	Equally-weighted				Value-weighted			
Fund Group	α_{Gross}	α_{Net}	**Russell**	**RSQ**	α_{Gross}	α_{Net}	**Russell**	**RSQ**
Large-cap Growth	0.12*	0.02	0.97	0.96	0.10	0.03	1.00	0.95
	(1.94)	(0.39)	(-1.49)		(1.28)	(0.33)	(-0.12)	
Large-cap Blend	0.01	-0.07***	0.96***	0.99	0.04	-0.03	0.96***	0.99
	(0.67)	(-3.19)	(-5.42)		(1.09)	(-0.93)	(-3.45)	
Large-cap Value	0.04	-0.05	0.93***	0.98	0.03	-0.03	0.94***	0.98
	(0.99)	(-1.23)	(-6.12)		(0.85)	(-0.77)	(-4.58)	
Mid-cap Growth	0.19***	0.08	0.91***	0.97	0.16**	0.07	0.93***	0.96
	(2.79)	(1.17)	(-4.05)		(2.38)	(1.02)	(-2.80)	
Mid-cap Blend	0.09**	-0.02	0.94***	0.98	0.07	-0.02	0.97*	0.97
	(2.10)	(-0.53)	(-4.89)		(1.18)	(-0.29)	(-1.86)	
Mid-cap Value	0.09	-0.01	0.88***	0.96	0.03	-0.05	0.90***	0.95
	(1.41)	(-0.23)	(-5.70)		(0.45)	(-0.65)	(-4.74)	
Small-cap Growth	0.40***	0.28***	0.90***	0.97	0.39***	0.29**	0.90***	0.97
	(5.28)	(3.65)	(-10.86)		(5.17)	(3.87)	(-7.94)	
Small-cap Blend	0.28***	0.17**	0.90***	0.96	0.24***	0.15**	0.90***	0.97
	(3.94)	(2.34)	(-4.48)		(3.73)	(2.33)	(-5.74)	
Small-cap Value	0.19***	0.08	0.92***	0.95	0.19***	0.10	0.92***	0.96
	(2.91)	(1.21)	(-6.60)		(2.88)	(1.52)	(-5.73)	

Table 4: Comparison of Schwarz Information Criteria of the Four Benchmark Models

The table shows values of the Schwarz Information Criterion (SIC) for the CAPM, three-factor (FF), the four-factor (FFM) and the Russell Indices (RS) versions of regression (1) estimated on the equal-weighted and the value-weighted style portfolios of funds. Figures in **bold** indicate the model with the lowest SIC. Estimates of coefficients of the various regressions are reported in Table 3.

	Equally-weighted				**Value-weighted**			
Fund Group	**CAPM**	**FF**	**FFM**	**RS**	**CAPM**	**FF**	**FFM**	**RS**
Large-cap Growth	3.164	2.548	**2.537**	2.748	3.583	**2.843**	2.847	3.231
Large-cap Blend	1.319	0.844	0.865	**0.680**	1.557	**1.431**	1.448	1.534
Large-cap Value	3.475	2.494	2.455	**1.916**	3.475	2.437	2.385	**1.898**
Mid-cap Growth	4.390	3.319	3.334	**2.889**	4.448	3.338	3.346	**3.056**
Mid-cap Blend	3.286	2.774	2.742	**2.030**	3.376	2.988	2.992	**2.677**
Mid-cap Value	3.925	3.222	**3.160**	3.399	4.173	3.442	**3.397**	3.755
Small-cap Growth	5.003	3.287	3.246	**2.875**	5.046	3.433	3.390	**2.961**
Small-cap Blend	4.508	3.251	3.053	**2.839**	4.568	3.225	3.047	**2.703**
Small-cap Value	4.589	3.424	3.197	**2.920**	4.612	3.451	3.193	**2.860**

Table 5A: Alpha Estimates from the Style-consistent Benchmarks and the Market Risk Factors (Net Returns)

This table reports average coefficients of regressions of individual fund returns on corresponding Russell style-consistent benchmarks (RS) and multi-factor risk measures – SMB, B/M and MOM. The coefficients are estimated for each fund in the given style group and then the cross-sectional mean and the t-statistic (in parenthesis) of the average coefficients are calculated. Panel A reports results for the actual coefficients. Panel B report comparable results, but where all independent regressors have been rescaled by their respective sample standard deviations (see text in Section 3.3 for information on the interpretation of the results in Panel B).

Panel A	Large Cap Growth		Large Cap Blend		Large Cap Value		Mid Cap Growth		Mid Cap Blend		Mid Cap Value		Small Cap Growth		Small Cap Blend		Small Cap Value	
Alpha	0.013	-0.104	-0.059	-0.090	-0.043	-0.033	0.022	-0.055	-0.071	-0.062	-0.066	-0.035	0.162	0.148	0.121	0.113	0.094	0.148
t(Alpha)	(0.83)	(-10.01)	(-7.33)	(-11.58)	(-3.23)	(-3.00)	(1.04)	(-2.97)	(-2.59)	(-2.66)	(-3.39)	(-2.58)	(5.26)	(5.59)	(4.53)	(4.67)	(3.97)	(7.20)
RS	0.979	0.995	0.974	0.970	0.941	0.945	0.907	0.916	0.943	0.924	0.900	0.903	0.902	0.939	0.926	0.960	0.918	0.924
t(RS)	(91.67)	(114.79)	(180.70)	(177.74)	(134.55)	(141.34)	(107.61)	(154.16)	(65.98)	(70.17)	(86.13)	(88.99)	(105.47)	(138.70)	(65.50)	(73.56)	(97.37)	(98.71)
SMB		0.130		0.054		0.054		0.076		0.054		0.017		-0.102		-0.093		-0.016
t(SMB)		(13.90)		(9.29)		(7.93)		(6.41)		(3.89)		(1.57)		(-11.47)		(-6.25)		(-1.07)
B/M		0.073		0.015		-0.114		0.071		0.006		-0.105		0.025		0.034		-0.133
t(B/M)		(6.29)		(2.23)		(-10.07)		(4.98)		(0.47)		(-9.70)		(1.43)		(1.34)		(-8.56)
MOM		0.103		0.013		-0.061		0.108		-0.007		-0.063		0.041		-0.012		-0.114
t(MOM)		(11.99)		(3.11)		(-9.46)		(9.21)		(-0.57)		(-7.15)		(2.94)		(-0.72)		(-13.47)
RSQ	0.871	0.900	0.913	0.930	0.887	0.909	0.847	0.879	0.843	0.871	0.868	0.890	0.867	0.890	0.865	0.890	0.855	0.874
LOGL	-288.2	-269.8	-236.3	-218.8	-266.2	-250.0	-352.8	-334.7	-318.0	-301.1	-267.1	-255.0	-342.3	-325.8	-283.9	-268.2	-321.7	-309.2

Panel B																		
Russell	4.940	5.021	4.324	4.305	4.070	4.089	5.589	5.646	4.658	4.567	4.239	4.254	6.112	6.364	5.320	5.513	4.687	4.715
t(Russell)	(91.67)	(114.79)	(180.70)	(177.74)	(134.55)	(141.34)	(107.61)	(154.16)	(65.98)	(70.17)	(86.13)	(88.99)	(105.47)	(138.70)	(65.50)	(73.56)	(97.37)	(98.71)
SMB		0.411		0.169		0.170		0.240		0.170		0.054		-0.321		-0.294		-0.049
t(SMB)		(13.90)		(9.29)		(7.93)		(6.41)		(3.89)		(1.57)		(-11.47)		(-6.25)		(-1.07)
B/M		0.289		0.058		-0.452		0.280		0.025		-0.416		0.098		0.132		-0.524
t(B/M)		(6.29)		(2.23)		(-10.07)		(4.98)		(0.47)		(-9.70)		(1.43)		(1.34)		(-8.56)
MOM		0.530		0.068		-0.316		0.560		-0.035		-0.328		0.214		-0.061		-0.586
t(MOM)		(11.99)		(3.11)		(-9.46)		(9.21)		(-0.57)		(-7.15)		(2.94)		(-0.72)		(-13.47)

Table 5B: Alpha Estimates from the Style-consistent Benchmarks and the Market Risk Factors (Gross Returns)

This table reports average coefficients of regressions of individual fund returns on corresponding Russell style-consistent benchmarks (RS) and multi-factor risk measures – SMB, B/M and MOM. The coefficients are estimated for each fund in the given style group and then the cross-sectional mean and the t-statistic (in parenthesis) of the average coefficients are calculated. Panel A reports results for the actual coefficients. Panel B report comparable results, but where all independent regressors have been rescaled by their respective sample standard deviations (see text in Section 3.3 for information on the interpretation of the results in Panel B).

Panel A	Large Cap Growth		Large Cap Blend		Large Cap Value		Mid Cap Growth		Mid Cap Blend		Mid Cap Value		Small Cap Growth		Small Cap Blend		Small Cap Value	
Alpha	0.110	-0.006	0.028	-0.003	0.049	0.059	0.130	0.055	0.044	0.052	0.035	0.065	0.287	0.273	0.230	0.222	0.205	0.259
t(Alpha)	(6.60)	(-0.56)	(3.57)	(-0.34)	(4.04)	(5.98)	(6.11)	(3.02)	(1.83)	(2.59)	(1.88)	(5.04)	(9.69)	(10.34)	(8.57)	(9.05)	(8.46)	(12.86)
RS	0.981	0.997	0.975	0.971	0.942	0.947	0.909	0.918	0.944	0.925	0.901	0.904	0.903	0.940	0.927	0.961	0.919	0.925
t(RS)	(91.188)	(114.59)	(180.63)	(177.70)	(134.24)	(140.37)	(108.64)	(155.94)	(65.94)	(70.18)	(86.11)	(89.02)	(105.32)	(139.41)	(65.50)	(73.51)	(97.38)	(98.75)
SMB		0.128		0.054		0.055		0.075		0.054		0.017		-0.101		-0.093		-0.016
t(SMB)		(13.73)		(9.20)		(7.81)		(6.41)		(3.91)		(1.58)		(-11.28)		(-6.24)		(-1.06)
B/M		0.073		0.014		-0.116		0.070		0.006		-0.105		0.025		0.034		-0.133
t(B/M)		(6.21)		(2.05)		(-9.92)		(4.85)		(0.46)		(-9.69)		(1.40)		(1.34)		(-8.56)
MOM		0.102		0.013		-0.062		0.107		-0.007		-0.063		0.041		-0.012		-0.114
t(MOM)		(11.63)		(2.95)		(-9.27)		(9.18)		(-0.57)		(-7.14)		(2.87)		(-0.72)		(-13.42)
RSQ	0.872	0.901	0.913	0.930	0.887	0.909	0.848	0.879	0.842	0.871	0.868	0.890	0.867	0.890	0.865	0.890	0.855	0.874
LOGL	-284.6	-266.6	-234.9	-217.6	-265.1	-248.9	-352.0	-334.1	-318.2	-301.2	-267.2	-255.1	-342.4	-326.0	-284.1	-268.4	-321.8	-309.4

Panel B																		
Russell	4.952	5.030	4.329	4.310	4.075	4.095	5.600	5.656	4.663	4.572	4.243	4.258	6.119	6.369	5.325	5.519	4.692	4.720
t(RS)	(91.188)	(114.59)	(180.63)	(177.70)	(134.24)	(140.37)	(108.64)	(155.94)	(65.94)	(70.18)	(86.11)	(89.02)	(105.32)	(139.41)	(65.50)	(73.51)	(97.38)	(98.75)
SMB		0.406		0.169		0.172		0.238		0.170		0.055		-0.320		-0.294		-0.049
t(SMB)		(13.73)		(9.20)		(7.81)		(6.41)		(3.91)		(1.58)		(-11.28)		(-6.24)		(-1.06)
B/M		0.287		0.054		-0.459		0.274		0.025		-0.416		0.097		0.133		-0.525
t(B/M)		(6.21)		(2.05)		(-9.92)		(4.85)		(0.46)		(-9.69)		(1.40)		(1.34)		(-8.56)
MOM		0.525		0.066		-0.321		0.552		-0.036		-0.328		0.211		-0.061		-0.587
t(MOM)		(11.63)		(2.95)		(-9.27)		(9.18)		(-0.57)		(-7.14)		(2.87)		(-0.72)		(-13.42)

Table 6: Distribution of Actual and Simulated Fund Performance (3 Factor and Russell models)

The table shows values of t(α) at selected percentiles (Pct.) of the distribution of t(α) estimates for actual **(Act)** gross and net fund returns. **Sim** is the average value of t(α) at the selected percentiles from the simulations. The table also shows the percentage of the 10,000 simulation runs that produce lower values of t(α) at the selected percentiles than those observed for actual fund returns **(% < Act)**. Statistically significant actuals (at 5% level of significance) are highlighted in bold print. See Figure 1 for related CDFs.

Panel A						
	3-Factor (gross)			**3-Factor (net)**		
Percentiles	**Actual**	**Sim**	**%<Actual**	**Actual**	**Sim**	**%<Actual**
1	**-2.49**	-2.37	32.4	**-3.43**	-2.37	0.4
2	**-2.04**	-2.08	51.5	**-2.88**	-2.08	1.6
3	-1.77	-1.90	63.7	**-2.60**	-1.90	2.5
4	-1.62	-1.76	65.5	**-2.40**	-1.77	3.3
5	-1.48	-1.65	70.2	**-2.27**	-1.66	3.5
10	-1.09	-1.28	74.1	-1.77	-1.29	6.2
90	**1.96**	1.28	98.0	1.30	1.28	56.3
95	**2.31**	1.65	97.5	1.66	1.65	54.6
96	**2.46**	1.76	97.8	1.78	1.76	56.8
97	**2.58**	1.89	97.5	1.88	1.89	51.9
98	**2.82**	2.08	98.1	**2.12**	2.07	59.6
99	**3.18**	2.37	98.5	**2.47**	2.37	65.3
Panel B						
	RS Benchmark (gross)			**RS Benchmark (net)**		
Percentiles	**Actual**	**Sim**	**%<Actual**	**Actual**	**Sim**	**%<Actual**
1	**-2.50**	-2.46	40.7	**-3.43**	-2.45	0.3
2	**-1.96**	-2.16	75.6	**-2.84**	-2.15	1.4
3	-1.72	-1.97	82.8	**-2.52**	-1.97	2.9
4	-1.51	-1.83	90.3	**-2.30**	-1.83	4.4
5	-1.33	-1.72	95.3	**-2.15**	-1.72	5.3
10	-0.85	-1.33	99.1	-1.66	-1.33	9.1
90	**2.18**	1.28	99.9	1.55	1.28	87.8
95	**2.70**	1.65	100.0	**2.02**	1.65	92.5
96	**2.81**	1.76	99.9	**2.17**	1.75	93.6
97	**2.98**	1.89	99.9	**2.32**	1.89	94.1
98	**3.21**	2.07	99.9	**2.52**	2.06	94.2
99	**3.62**	2.35	99.9	**2.92**	2.35	96.3

Table 7: Distribution of Actual and Simulated Net Fund Performance by Style (Three-Factor Benchmark)

(Act) actual net fund returns t(α). (Sim) is the average value of the simulations t(α), (% < Act) the percentage of simulations runs that produce lower values of t(α). Statistically significant actuals (at 5% level of significance) are highlighted in bold print. See Figure 2 for related CDFs.

	Large Growth			Large Blend			Large Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	**-3.41**	-2.33	1.9	**-4.73**	-2.37	0.0	**-3.08**	-2.22	3.7
2	**-2.82**	-2.04	5.2	**-3.59**	-2.08	0.0	**-2.72**	-1.94	4.5
3	**-2.61**	-1.87	5.7	**-3.31**	-1.90	0.0	**-2.54**	-1.76	4.4
4	**-2.42**	-1.74	6.7	**-3.05**	-1.76	0.1	**-2.38**	-1.63	4.8
5	**-2.36**	-1.63	5.5	**-2.85**	-1.65	0.1	**-2.23**	-1.53	5.6
10	-1.83	-1.27	9.2	**-2.27**	-1.28	0.2	-1.69	-1.17	11.2
90	0.88	1.19	22.7	0.84	1.29	4.6	1.31	1.25	58.8
95	1.21	1.54	21.9	1.35	1.66	14.5	1.67	1.61	58.1
96	1.27	1.64	18.5	1.44	1.77	12.7	1.81	1.72	61.2
97	1.32	1.77	12.9	1.52	1.91	9.7	1.91	1.86	57.9
98	1.53	1.94	16.2	1.62	2.09	5.8	1.97	2.05	46.1
99	1.80	2.21	17.0	**2.07**	2.39	18.8	**2.34**	2.36	51.2
	Mid-Cap Growth			Mid-Cap Blend			Mid-Cap Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	**-2.11**	-2.24	57.6	**-3.86**	-2.26	0.3	**-2.86**	-2.25	13.4
2	**-1.96**	-1.96	47.7	**-2.85**	-1.98	4.0	**-2.63**	-1.95	8.8
3	-1.83	-1.79	43.8	**-2.65**	-1.81	3.8	**-2.31**	-1.77	13.3
4	-1.69	-1.66	44.9	**-2.52**	-1.68	3.7	**-2.12**	-1.64	15.4
5	-1.64	-1.56	40.8	**-2.35**	-1.57	4.5	-1.89	-1.53	21.2
10	-1.30	-1.21	40.2	-1.56	-1.21	18.9	-1.19	-1.17	46.8
90	1.39	1.17	70.1	1.75	1.24	88.3	1.75	1.24	86.4
95	1.73	1.51	69.1	**2.12**	1.59	88.7	**2.02**	1.59	82.4
96	1.79	1.62	66.4	**2.27**	1.70	90.0	**2.17**	1.69	84.7
97	1.95	1.74	69.1	**2.36**	1.83	88.2	**2.53**	1.82	92.2
98	**2.25**	1.91	77.1	**2.40**	2.01	81.6	**2.88**	2.00	95.2
99	**2.49**	2.18	74.9	**3.26**	2.28	96.1	**3.31**	2.27	96.1
	Small Growth			Small Blend			Small Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	**-3.11**	-2.28	6.1	**-3.58**	-2.25	0.7	**-2.24**	-2.12	37.8
2	**-2.74**	-2.00	6.7	**-2.65**	-1.95	7.3	**-2.01**	-1.85	34.6
3	**-2.39**	-1.82	11.2	**-2.59**	-1.77	4.5	-1.85	-1.68	33.0
4	**-2.28**	-1.69	10.5	**-2.40**	-1.64	5.5	-1.71	-1.56	34.7
5	**-2.14**	-1.59	11.5	**-2.22**	-1.53	7.0	-1.61	-1.46	34.9
10	-1.74	-1.23	12.8	-1.74	-1.18	10.7	-0.96	-1.11	60.5
90	1.14	1.21	44.5	1.48	1.22	74.0	1.64	1.22	81.2
95	1.75	1.55	68.9	1.70	1.57	64.1	1.88	1.58	74.6
96	1.80	1.65	64.8	1.82	1.68	65.5	**2.10**	1.69	80.4
97	1.85	1.78	58.1	**2.21**	1.81	82.0	**2.30**	1.82	83.1
98	1.91	1.95	48.5	**2.40**	1.98	81.6	**2.77**	2.00	91.8
99	**2.30**	2.22	58.7	**2.76**	2.27	84.0	**2.97**	2.28	88.1

Table 8: Distribution of Actual and Simulated Net Fund Performance by Style (Russell Index Benchmark)

(Act) actual net fund returns t(α). (Sim) is the average value of the simulations t(α), (% < Act) the percentage of simulations runs that produce lower values of t(α). Statistically significant actuals (at 5% level of significance) are highlighted in bold print. See Figure 3 for related CDFs.

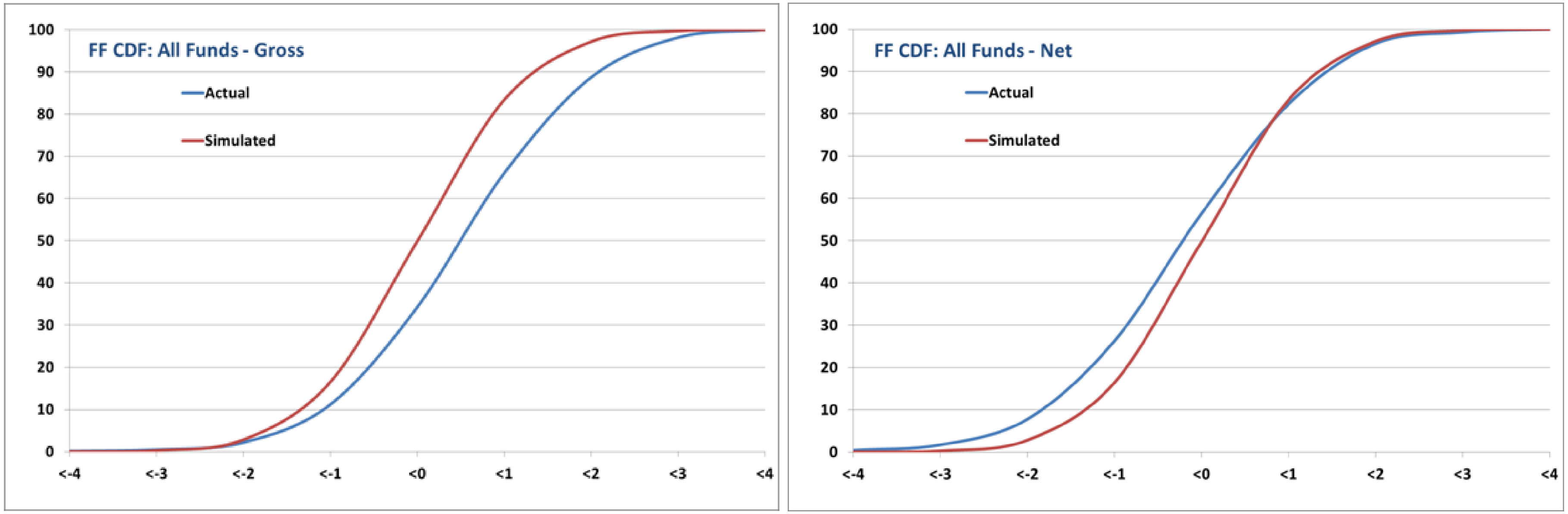
	Large Growth			Large Blend			Large Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	**-2.69**	-2.29	17.8	**-4.44**	-2.37	0.0	**-2.92**	-2.29	9.3
2	**-2.30**	-2.01	23.3	**-3.73**	-2.08	0.0	**-2.71**	-2.02	6.4
3	**-2.12**	-1.84	23.2	**-3.31**	-1.91	0.0	**-2.52**	-1.85	6.1
4	**-2.06**	-1.71	19.1	**-3.06**	-1.77	0.0	**-2.48**	-1.72	3.9
5	**-1.97**	-1.61	17.4	**-2.84**	-1.66	0.1	**-2.29**	-1.61	5.3
10	-1.45	-1.25	28.7	**-2.19**	-1.30	0.2	-1.91	-1.25	5.4
90	1.39	1.21	70.1	0.97	1.29	9.4	1.13	1.24	40.4
95	1.54	1.56	51.8	1.23	1.65	4.8	1.62	1.59	55.2
96	1.56	1.66	43.1	1.44	1.76	12.3	1.78	1.70	61.4
97	1.61	1.78	35.4	1.57	1.89	13.3	1.90	1.83	59.5
98	1.73	1.95	31.6	1.75	2.06	14.6	**2.05**	2.01	57.0
99	**2.13**	2.21	44.7	**2.27**	2.35	43.9	**2.52**	2.31	71.0

	Mid-Cap Growth			Mid-Cap Blend			Mid-Cap Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	**-2.44**	-2.31	36.4	**-3.44**	-2.35	1.7	**-3.20**	-2.90	28.0
2	**-2.11**	-2.03	40.2	**-3.28**	-2.07	0.4	**-2.98**	-2.57	20.1
3	-1.78	-1.85	55.4	**-3.06**	-1.89	0.3	**-2.76**	-2.36	19.9
4	-1.63	-1.72	57.4	**-2.93**	-1.76	0.2	**-2.20**	-2.21	48.6
5	-1.54	-1.62	55.5	**-2.57**	-1.65	0.9	**-2.14**	-2.09	43.4
10	-1.22	-1.26	52.8	**-2.12**	-1.28	0.6	-1.78	-1.68	39.0
90	1.47	1.23	75.8	1.42	1.27	69.4	1.15	1.12	56.1
95	1.82	1.57	75.1	1.85	1.63	75.5	1.53	1.51	54.4
96	1.91	1.68	74.6	1.92	1.74	71.4	1.78	1.63	66.8
97	1.95	1.80	67.4	**2.08**	1.87	73.5	1.95	1.77	68.5
98	**2.31**	1.97	80.6	**2.19**	2.04	67.5	**2.23**	1.96	74.5
99	**2.76**	2.23	88.4	**2.36**	2.31	57.1	**2.52**	2.26	71.9

	Small Growth			Small Blend			Small Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	**-2.12**	-2.32	65.0	**-5.25**	-2.30	0.0	**-2.05**	-2.17	56.8
2	-1.69	-2.04	79.6	**-4.36**	-2.01	0.0	-1.80	-1.91	56.6
3	-1.58	-1.86	75.0	**-3.02**	-1.84	0.7	-1.62	-1.74	58.1
4	-1.54	-1.73	67.4	-1.94	-1.71	26.9	-1.34	-1.62	70.8
5	-1.33	-1.63	77.0	-1.78	-1.60	30.5	-1.21	-1.52	73.5
10	-0.84	-1.26	88.6	-1.28	-1.24	43.7	-0.94	-1.18	68.0
90	**2.33**	1.24	99.8	**2.15**	1.23	98.5	**2.06**	1.17	96.3
95	**2.86**	1.59	99.9	**2.47**	1.57	97.9	**2.39**	1.50	96.0
96	**2.97**	1.69	99.9	**2.57**	1.68	97.7	**2.56**	1.60	96.7
97	**3.07**	1.82	99.7	**2.69**	1.80	97.4	**2.80**	1.72	97.8
98	**3.26**	1.99	99.7	**2.89**	1.97	97.4	**2.95**	1.88	97.5
99	**3.53**	2.26	99.5	**3.12**	2.25	95.8	**3.22**	2.14	96.7

Figure 1: Cumulative Density Functions of t(alphas) for Gross and Net Returns

Panel A: Fama and French results (see Table 5, Panel A)



Panel A: RS Benchmark results (see Table 5, Panel B)

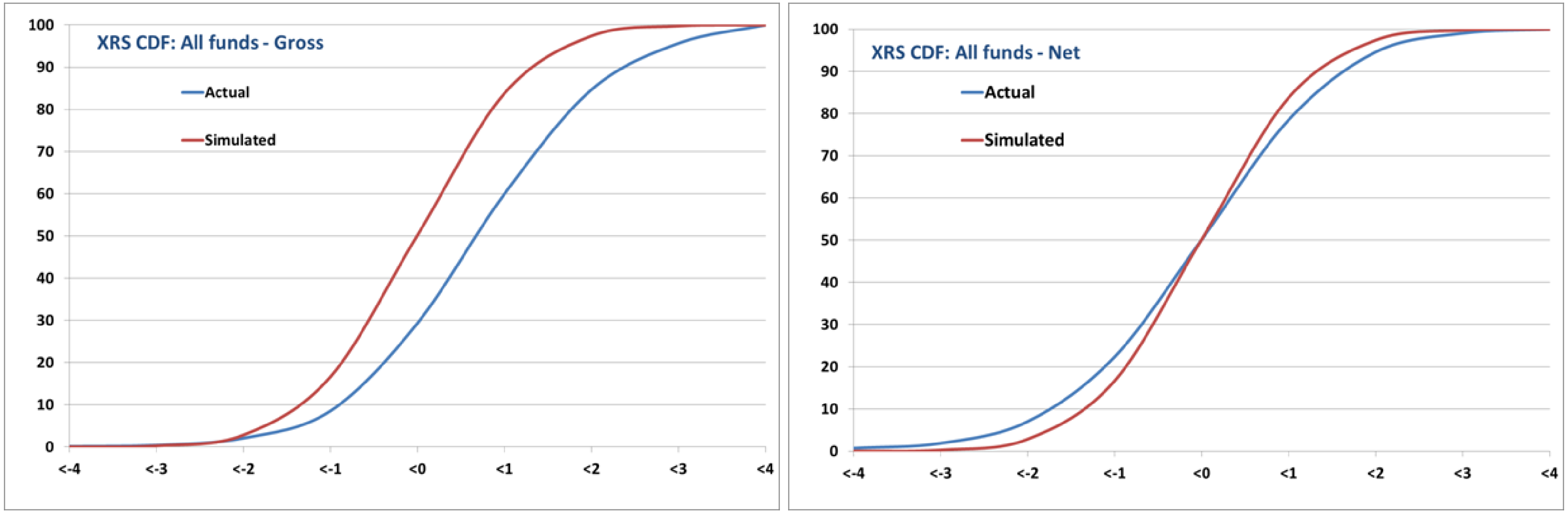


Figure 2: Cumulative Density Functions based on 3-Factor t(alphas) for Net Returns by Style (See Table 6)

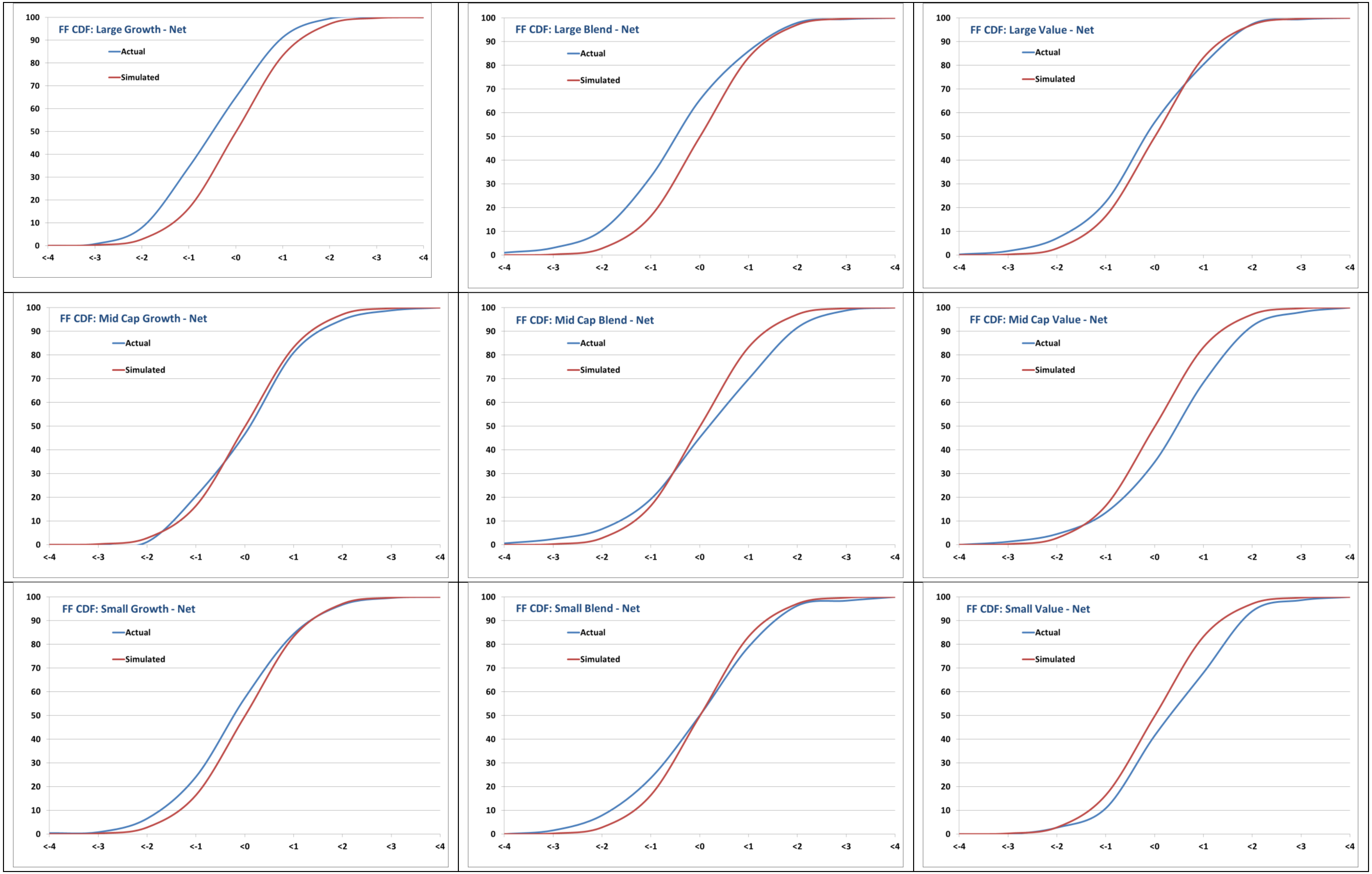
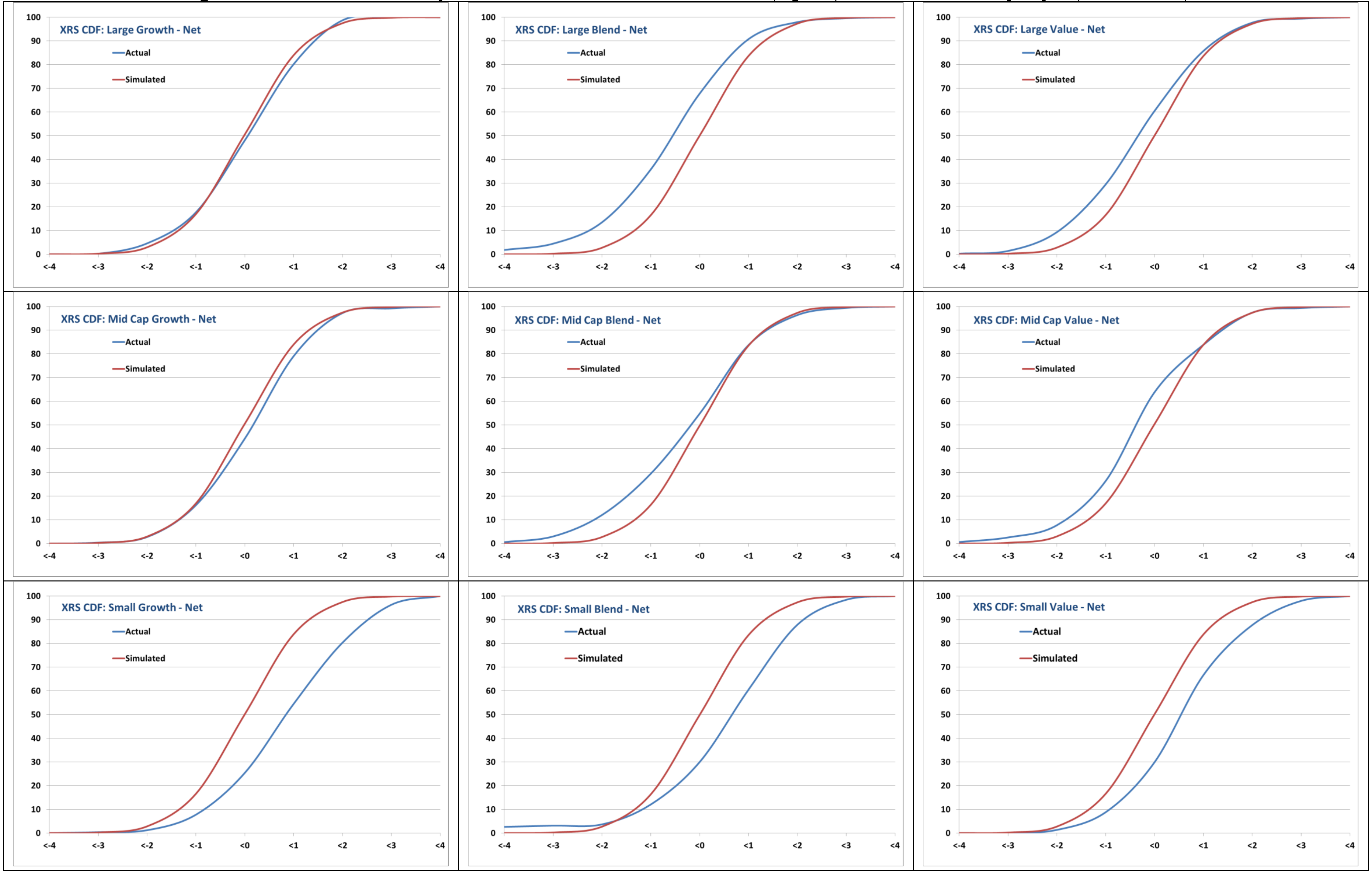


Figure 3: Cumulative Density Functions based on Russell Index t(alphas) for Net Returns by Style (See Table 7)



Appendix 1: Distribution of Actual and Simulated Fund Performance (CAPM and 4 Factor models)

The table shows values of t(α) at selected percentiles (Pct.) of the distribution of t(α) estimates for actual **(Act)** gross and net fund returns. **Sim** is the average value of t(α) at the selected percentiles from the simulations. The table also shows the percentage of the 10,000 simulation runs that produce lower values of t(α) at the selected percentiles than those observed for actual fund returns **(% < Act)**. Statistically significant actuals (at 5% level of significance) are highlighted in bold print.

Panel A

	CAPM (gross)			CAPM (net)		
Percentiles	**Actual**	**Sim**	**%<Actual**	**Actual**	**Sim**	**%<Actual**
1	**-2.40**	**-2.31**	35.4	**-3.19**	**-2.31**	2.7
2	**-1.98**	**-2.04**	50.5	**-2.81**	**-2.04**	3.8
3	**-1.66**	**-1.87**	68.2	**-2.46**	**-1.87**	6.9
4	-1.51	**-1.74**	72.3	**-2.25**	**-1.74**	9.2
5	-1.38	-1.64	75.0	**-2.09**	-1.64	11.1
10	-0.89	-1.28	90.2	-1.56	-1.28	18.7
90	**1.97**	1.26	97.2	1.52	1.26	80.6
95	**2.43**	1.60	98.1	**1.92**	1.61	83.2
96	**2.56**	**1.70**	98.3	**2.03**	**1.71**	83.4
97	**2.72**	**1.83**	98.5	**2.24**	**1.83**	87.8
98	**2.94**	**1.99**	98.7	**2.42**	**2.00**	88.0
99	**3.28**	**2.26**	99.0	**2.79**	**2.27**	91.4

Panel B

	4 Factor (gross)			4 Factor (net)		
Percentiles	**Actual**	**Sim**	**%<Actual**	**Actual**	**Sim**	**%<Actual**
1	**-2.80**	**-2.39**	10.8	**-3.57**	**-2.39**	0.2
2	**-2.43**	**-2.09**	13.5	**-3.19**	**-2.10**	0.3
3	**-2.06**	**-1.91**	28.5	**-2.98**	**-1.91**	0.3
4	**-1.87**	**-1.77**	33.5	**-2.74**	**-1.77**	0.5
5	**-1.71**	**-1.66**	39.4	**-2.54**	**-1.66**	0.7
10	-1.26	-1.29	49.5	**-2.06**	-1.29	0.9
90	1.73	1.30	92.6	1.10	1.29	27.4
95	**2.17**	**1.67**	94.1	1.51	**1.66**	33.8
96	**2.29**	**1.78**	94.1	1.64	**1.77**	36.3
97	**2.49**	**1.91**	95.4	**1.82**	**1.91**	42.0
98	**2.71**	**2.10**	95.8	**1.97**	**2.09**	38.4
99	**3.04**	**2.39**	96.2	**2.36**	**2.38**	51.1

Appendix 2: Distribution of Actual and Simulated Net Fund Performance by Style (CAPM)

(Act) actual net fund returns t(α). (Sim) is the average value of the simulations t(α), (% < Act) the percentage of simulations runs that produce lower values of t(α). Statistically significant actuals (at 5% level of significance) are highlighted in bold print.

	Large Growth			Large Blend			Large Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	-3.33	-2.19	2.0	-4.01	-2.31	0.0	-3.07	-2.09	3.8
2	-3.05	-1.92	1.9	-3.63	-2.02	0.0	-2.46	-1.82	10.1
3	-2.82	-1.75	2.2	-3.20	-1.85	0.1	-2.30	-1.65	10.0
4	-2.58	-1.63	3.6	-2.92	-1.71	0.2	-2.14	-1.53	11.4
5	-2.42	-1.53	4.6	-2.78	-1.61	0.2	-1.96	-1.43	13.9
10	-1.92	-1.18	7.8	-2.18	-1.25	1.0	-1.43	-1.10	23.4
90	0.70	1.15	17.3	0.75	1.27	3.3	1.04	1.13	47.1
95	0.94	1.48	11.2	1.10	1.62	3.5	1.44	1.47	51.8
96	1.13	1.58	16.6	1.29	1.73	7.7	1.56	1.57	53.2
97	1.21	1.70	14.0	1.43	1.86	8.5	1.63	1.70	48.3
98	1.39	1.86	15.4	1.59	2.03	8.5	1.82	1.87	50.3
99	1.71	2.12	20.3	2.06	2.32	25.4	2.07	2.16	46.8

	Mid-Cap Growth			Mid-Cap Blend			Mid-Cap Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	-2.48	-2.06	23.3	-2.87	-2.17	10.0	-1.97	-2.13	58.6
2	-1.84	-1.80	44.4	-2.57	-1.91	9.7	-1.64	-1.87	63.6
3	-1.42	-1.64	62.0	-2.39	-1.74	9.8	-1.48	-1.70	63.0
4	-1.30	-1.53	62.5	-2.13	-1.62	14.7	-1.43	-1.57	57.8
5	-1.26	-1.43	58.7	-1.94	-1.52	18.6	-1.33	-1.47	58.0
10	-0.92	-1.11	60.9	-1.15	-1.17	49.4	-0.83	-1.14	70.7
90	1.58	1.07	82.0	2.01	1.19	95.2	1.88	1.15	92.0
95	1.94	1.38	84.3	2.33	1.52	94.7	2.29	1.46	93.9
96	2.30	1.47	93.1	2.50	1.62	95.9	2.48	1.56	95.7
97	2.40	1.59	92.9	2.72	1.74	97.1	2.77	1.67	97.4
98	2.64	1.74	94.6	2.82	1.90	96.1	2.91	1.83	97.1
99	2.74	1.98	91.0	3.21	2.16	96.9	3.08	2.07	96.0

	Small Growth			Small Blend			Small Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	-1.92	-1.97	51.1	-2.14	-2.00	38.9	-1.99	-1.89	42.0
2	-1.36	-1.73	70.0	-1.68	-1.74	51.4	-1.24	-1.65	72.8
3	-1.25	-1.57	68.0	-1.53	-1.58	51.2	-1.08	-1.50	72.7
4	-1.16	-1.46	66.4	-1.41	-1.47	51.6	-0.94	-1.39	74.6
5	-1.09	-1.37	65.4	-1.20	-1.37	59.2	-0.66	-1.30	83.4
10	-0.70	-1.07	70.6	-0.47	-1.06	83.0	-0.26	-1.00	86.9
90	1.76	1.01	88.4	1.94	1.03	92.7	2.07	0.99	94.3
95	2.07	1.30	89.1	2.34	1.33	94.8	2.56	1.29	97.0
96	2.20	1.39	90.3	2.36	1.42	93.6	2.64	1.37	96.9
97	2.30	1.49	90.3	2.53	1.53	94.8	2.82	1.48	97.7
98	2.37	1.63	88.6	2.63	1.67	94.1	2.87	1.63	96.7
99	2.95	1.86	95.2	2.81	1.91	93.0	2.92	1.86	94.2

Appendix 3: Distribution of Actual and Simulated Net Fund Performance by Style (4 Factor)

(Act) actual net fund returns t(α). (Sim) is the average value of the simulations t(α), (% < Act) the percentage of simulations runs that produce lower values of t(α). Statistically significant actuals (at 5% level of significance) are highlighted in bold print.

	Large Growth			Large Blend			Large Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	-3.17	-2.35	4.6	-4.67	-2.39	0.0	-3.29	-2.26	2.0
2	-2.91	-2.06	3.4	-3.72	-2.09	0.0	-2.83	-1.97	3.3
3	-2.51	-1.88	7.8	-3.28	-1.91	0.0	-2.78	-1.80	1.8
4	-2.33	-1.75	8.6	-3.08	-1.77	0.0	-2.62	-1.67	2.1
5	-2.01	-1.64	18.0	-2.96	-1.66	0.0	-2.50	-1.56	2.0
10	-1.69	-1.28	15.3	-2.38	-1.29	0.1	-2.10	-1.20	2.4
90	1.09	1.19	41.6	0.92	1.29	9.1	0.86	1.26	16.0
95	1.37	1.55	34.6	1.29	1.66	10.3	1.20	1.63	14.1
96	1.50	1.65	38.0	1.45	1.77	14.5	1.30	1.74	14.1
97	1.59	1.78	33.9	1.51	1.91	9.5	1.37	1.87	10.6
98	1.71	1.96	29.9	1.68	2.10	9.3	1.68	2.05	19.4
99	1.89	2.24	21.5	2.03	2.40	15.6	2.19	2.35	38.7

	Mid-Cap Growth			Mid-Cap Blend			Mid-Cap Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	-2.50	-2.26	28.8	-3.70	-2.28	0.7	-3.18	-2.30	7.5
2	-2.07	-1.97	39.5	-2.69	-1.99	6.9	-2.58	-1.98	12.1
3	-1.89	-1.80	39.9	-2.48	-1.82	7.0	-2.46	-1.79	9.4
4	-1.66	-1.67	48.1	-2.43	-1.69	5.1	-2.29	-1.66	10.0
5	-1.62	-1.56	43.2	-2.31	-1.58	5.1	-2.00	-1.55	16.2
10	-1.41	-1.21	32.8	-1.90	-1.22	5.9	-1.43	-1.19	28.4
90	1.34	1.18	65.6	1.50	1.24	74.2	1.55	1.24	76.0
95	1.69	1.53	66.1	1.94	1.60	79.2	1.91	1.60	76.0
96	1.83	1.63	68.9	1.97	1.71	74.4	2.01	1.70	75.5
97	1.98	1.76	70.0	2.06	1.84	71.1	2.36	1.83	86.1
98	2.12	1.93	67.3	2.20	2.02	67.8	2.74	2.01	92.0
99	2.54	2.21	74.9	2.46	2.30	65.7	3.22	2.28	94.3

	Small Growth			Small Blend			Small Value		
Percentiles	Actual	Sim	%<Actual	Actual	Sim	%<Actual	Actual	Sim	%<Actual
1	-3.19	-2.29	4.5	-5.11	-2.28	0.0	-3.39	-2.22	2.6
2	-3.10	-2.00	1.7	-3.97	-1.99	0.0	-3.06	-1.94	1.8
3	-3.02	-1.83	1.0	-3.65	-1.81	0.0	-2.66	-1.76	3.9
4	-2.71	-1.69	1.7	-3.59	-1.67	0.0	-2.58	-1.63	3.0
5	-2.61	-1.59	1.6	-3.44	-1.57	0.0	-2.31	-1.53	5.4
10	-2.13	-1.23	2.3	-3.05	-1.21	0.0	-2.06	-1.17	3.4
90	0.87	1.23	20.2	1.23	1.27	49.1	1.02	1.25	31.1
95	1.36	1.58	30.7	1.65	1.63	55.4	1.79	1.61	67.5
96	1.42	1.68	27.8	1.76	1.73	55.3	1.83	1.72	62.9
97	1.54	1.81	27.6	1.90	1.86	56.0	2.05	1.85	69.2
98	1.85	1.98	40.0	2.03	2.04	51.9	2.46	2.03	82.0
99	2.25	2.26	51.0	2.64	2.33	75.1	2.82	2.30	83.5