Neglected risks in mutual fund performance measurement: An additional cost to stock-picking.

Justus Heuer *

Version 1 - November 2012

Abstract

This paper takes a closer look at utility based performance measurement proposed by Goetzmann et al. (2007) and used in popular Morningstar star ratings. Utility based performance measures offer a very intuitive way of risk correction and are hard to manipulate. They require, however, a proper benchmark measure to filter out lucky funds. I propose to use the Daniel et al. (1997) (DGTW) characteristic based benchmark portfolios as benchmarks for the utility based performance measure. I find that the DGTW selection measure consistently overestimates the manager's selection skills in certainty equivalent terms, and that this overestimation can be decomposed into an idiosyncratic and a systematic component. In diversified fund portfolios, the certainty equivalent selection measure is, on average, 87*bps* higher than in undiversified fund portfolios. The remaining undiversifiable risks cost 47*bps* per year in certainty equivalent terms and can be explained in part be imprecise correction for known systematic risk factors, in part by unknown but undiversifiable risk factors. The certainty equivalent measure captures risk particularly well in years with high moment realizations of the CRSP value weighted index.

EFM Classification: 330, 370, 380

*Justus Heuer: CDSB and Lehrstuhl für ABWL, insbesondere Bankbetriebslehre. Universität Mannheim, L5, 2 am Schloss, 68161 Mannheim. `jheuer[at]mail.uni-mannheim.de`; tel. +49 621 181 2355.

1 Introduction

The analysis of mutual fund performance in the literature, as well as the fund selection process of the investor, has to rely on ex-post returns of the different investment portfolios. In an ex-post analysis, any of these observed returns can be the result of (1) market reward for systematic risks[1] taken by the manager, (2) the manager's investment skill or (3) mere luck. A proper performance measure should be able to differentiate between returns that have been generated by loading on priced, systematic risk factors or by luck, and those that are the result of true skill, as only skill can be persistent and thus justify the high fees on active management[2]. In addition, the risk averse investor will want to avoid bearing *undiversified* and *undiversifiable* idiosyncratic risk. The first can be found in funds that are not properly diversified, maybe because they pursue a specific strategy or focus on a certain sector. It can be eliminated by investing in several different funds. The latter refers to risk that naturally arises when deviating from the market portfolio and is characteristic of active management[3]. Minimizing such *undiversifiable idiosyncratic* risk is, along with the identification of mispriced assets, a key responsibility of investment management[4]. Truely *undiversifiable idiosyncratic* risk has to be absorbed by the investor and thus carries the same price tag as systematic risk. A skilled manager will only generate excess utility to the investor if the excess returns from his active portfolio exceed the price of the additional *undiversifiable idiosyncratic* risk. In a performance model, it can therefore be treated just like systematic risks.

The most widely used performance measure in the literature, the Carhart (1997) alpha, uses a one shot procedure to correct for systematic risks and *characteristic luck*[5]. However, in

[1] *Systematic risk* will refer to any risks —- of any order — which are rewarded. The discussion of systematic risks reduces to the CAPM risk return relationship if the returns are normally distributed or the investor follows quadratic utility.

[2] Cremers and Petajisto (2009) find that funds charge on average 124*bps* per year, or 89*bps* value weighted. In our, newer sample, fees are on average 121*bps*

[3] *Undiversifiable idiosyncratic* risk is conceptually similar to the tracking error with the distinction that *undiversifiable idiosyncratic* allows for a more sophisticated benchmark concept. It can be characterized by the residual variance of a regression of portfolio returns on systematic risk premiums. While the residual variance might not be priced, it cannot be diversified away if there is only a limited number of profitable active portfolios (mispriced assets).

[4] Treynor and Black (1973) propose to hold a weighted combination of the active portfolio and the market portfolio to minimize such idiosyncratic risk.

[5] *Characteristic luck* will refer to lucky, i.e. unexpected superior performance of the characteristic (size, value

the context discussed above, the Carhart (1997) measure has two shortcomings. First, it fails to correct for *undiversifiable idiosyncratic* risk, which is left in the residuals. This makes it a useful measure to answer the question if fund managers can identify mispriced assets, it does not provide information if the excess returns come at the cost of poor diversification and thus substantial idiosyncratic but undiversifiable risk. Hence, a positive Carhart alpha portfolio can have a risk-return relationship, as measured e.g. by the Sharpe ratio, that is well below that of the market portfolio. The additional utility to the investor of such a portfolio is limited at best[6]. Second, it proxies for higher order systematic risk, or systematic risk orthogonal to that of the market, by using the Fama & French factors and a momentum factor. This is at best controversial[7]. Goetzmann et al. (2007) show that alpha measures can be manipulated. They propose a utility based performance model that treats all risks identically. Further, their Manipulation Proof Performance Measure (MPPM) does not rely on assumptions regarding the dimensionality of systematic risks and the distribution of the returns. Benchmarked against a market index, it allows to quantify the investor's certainty equivalent excess returns received from the fund manager's active management in total. It does, however, not allow to conclude if these returns are the result of luck or skill. While — looking at ex-post returns — it is impossible to tell with absolute certainty if returns are the result of luck or skill on an individual fund level, we can at least increase our chances by using a more sophisticated benchmark than a simple market index for the MPPM. The most important property of such a benchmark should be, that returns of the benchmark strongly covary with those of the fund's assets. The benchmark should be lucky whenever the fund is.

I propose to use the 125 size-, value-, and momentum sorted benchmark portfolios of Daniel et al. (1997) (DGTW) as the luck correction system for the MPPM. According to Daniel and Titman (1997), firms within any of these portfolios have similar properties and covary with one another. Thus, they span a very detailed definition of the market environment for any asset in the same

and momentum) strategy of the fund.

[6]The four factor alpha information ratio would be a way to address this problem.

[7]Daniel and Titman (1997) disputes that SMB and HML are factors explaining future consumption growth are orthogonal to the market and Chung et al. (2006) show that Fama & French Factors are proxies for higher order systematic risks. Liew and Vassalou (2000) show that the three factors do contain significant information about future GDP growth that is independent of the market factor.

portfolio. Like the Carhart measure, but more precisely so because the benchmark can change over time, the portfolios do correct for characteristic luck. Using this DGTW characteristic selectivity measure, but comparing MPPM certainty equivalents instead of plain returns, the certainty equivalent excess returns generated by the manager's stock picking skills can be identified. In addition, the manager needs to be credited for his predictions of the characteristic market environment. Again, the DGTW characteristic timing measure will be calculated as a difference in certainty equivalents. This allows to test whether, on a risk adjusted basis, managers move assets into superior performing characteristics from period to period. Timing performance that was not preceded by reallocations is not considered to be predicted by fund manager and is not attributed to his timing skill. It is considered *characteristic luck*.

Given that the size, book-to-market and momentum quintiles do not completely characterize an asset's risk profile, the plain Daniel et al. (1997) measures do not provide a proper risk correction. While this might be only of temporary importance for their timing measure (an asset manager cannot systematically trade towards a riskier portfolio every period, as he will hold the riskiest of all portfolios after a maximum of 125 periods), it is a serious caveat for the selection measure. For a manager systematically holding the riskiest out of all assets in any of the benchmark portfolios, the DGTW measure will falsely signal selection skills. In light of the ambiguous results on the systematic-risk-story related to the sorting factors, this lack of risk correction becomes even more severe[8]. In line with this logic, this paper shows that the simple DGTW selection measure consistently overestimates the manager's stock picking skills compared to the certainty equivalent MPPM measure[9]. Further, it is shown that the timing measure, as expected, does not require an additional risk correction.

I hypothesize that the overestimation of the stock picking skill has four sources, some of which can be eliminated by holding wider fund portfolios, others are undiversifiable: First, incomplete correction for the market risk and — to lesser extend — the factors size, value and momentum

[8] If the sorting factors largely characterize the systematic risk of an asset, than the discrepancy of systematic risk within any portfolio would be small, leaving the error in the selection measure small. If, however, there is any other determinant of systematic risk that is orthogonal to the sorting factors, the discrepancy in systematic risk within any portfolio and hence the bias can be potentially large.

[9] I will refer the difference between the new certainty equivalent based measure and the old DGTW measure as *selection spread* for the rest of the paper.

should drive the selection spread. This relation should be robust to diversification. Second, higher order systematic risk factors will explain some of the spread and are undiversifiable. While the exact nature of these factors remains opaque, the proposed methodology can at least quantify their costs. Third, undiversified idiosyncratic risks on an individual fund level should explain parts of the difference. These can be eliminated by investing in broader fund portfolios. The spread should decrease with the number of funds in the portfolio. Last, herding might explain part of the spread. Theoretically idiosyncratic risk becomes undiversifiable for a mutual fund investor if all mutual funds identify the same allegedly mispriced assets. To the mutual fund investor, herding has a cost because it systematizes idiosyncratic risk.

In line with O'Neal (1997), the analysis of multi-fund portfolios suggests that equity mutual fund investors should hold at least five different funds to be properly diversified. The selection spread decreases rapidly from on average 1.32% in one fund portfolios to 0.62% in five fund portfolios. Further diversification, however, only leads to slight reductions. The spread still remains significant ($p < 0.01$) 0.47% p.a. in 30 fund portfolios. Thus, the DGTW selection measure overestimates the return on manager's stock picking skills by almost 0.5% p.a. even in diversified portfolios. Imprecise adjustment for factor risks can explain most of this spread in a one factor model. In a four factor model, known risks are even overestimated by the DGTW measure. In this model, the spread is entirely due to unobserved risk factors or limits to diversification. Therefore, in addition to trading costs and mutual fund fees, higher undiversifiable risk exposure of mutual fund portfolios, compared to passive benchmark portfolios, further reduces the returns of stock picking by almost half a percentage.

The well regarded Morningstar star rating[10] relies on a utility based performance measure almost identical to the MPPM. Contrary to Goetzmann et al. (2007), Morningstar use all other funds of the same style category as benchmarks for each fund. While this might be a suitable characteristic luck correction, it disregards one important option investors have: passive investment. Further, it only allows to compare funds within a certain category. Investors probably

[10] Morningstar (2007) provide a detailed description of their rating methodology. Goetzmann et al. (2007) points out the similarity to his manipulation proof performance measure. Del Guercio and Tkac (2008) find a significant influence of Morningstar Ratings on mutual fund flows.

care more about their global performance, than about their relative performance conditional on investing into some category.

This paper's main contribution is suggesting to use characteristic sorted portfolios as future benchmarks in utility based performance measurement, as opposed to using a simple market index or other, comparable funds. Another main contribution is the quantification of additional risk-costs that are associated with stock picking. Specifically, I find that the traditional DGTW stock picking measure overestimates returns to stock picking by 47 in certainty equivalent terms. Further, I believe this is the first paper to discuss the costs of mutual fund herding to investors, caused by its implications for portfolio diversification. Finally, prior results regarding the necessity to hold diversified fund portfolios and DGTW results regarding market timing of mutual fund managers are confirmed.

The papers closest to this are Daniel et al. (1997) and Wermers (2000), while it also boroughs the main concept from Goetzmann et al. (2007). It further builds on ideas from Carhart (1997) and Cremers and Petajisto (2009). Interesting discussions of luck and skill in mutual fund performance can be found in Fama and French (2010) and Barras et al. (2010), these papers, however, focus on the cross section and do not look at funds individually.

The remainder of this paper proceeds as follows: Section 2 gives a (limited) overview of the literature on performance measurement. Section 3 introduces the new performance measure. Section 4 describes the data. Section 5 provides the results along with a detailed analysis of the source of the selection spread and section 6 concludes.

2 Literature

I identify two strands of literature that I aim to combine. The first focuses on the risk correction of the realized returns, the second deals with the manager's skill set. Of course, both strands are closely related as only risk adjusted returns are skilled returns. I provide only a brief overview of measures closely related to the proposed methodology. For a more complete discussion refer to Aragon and Ferson (2007).

I distinguish between systematic and total risk measures, which treat systematic and idiosyncratic risk equivalently. The most well known total risk correction is the Sharpe (1966) ratio, the excess portfolio return over the portfolio volatility. This ratio, however, strongly relies on the assumption of normality (or quadratic utility). The most common systematic risk correction is the Jensen (1969) alpha, the intercept of a regression of the portfolio return on the market return. Unfortunately, it relies on the same normality assumption as the Sharpe ratio and therefore only corrects for first order systematic risk. As a result of their strict normality assumption, Goetzmann et al. (2007) discuss how these measures can be manipulated and Agarwal et al. (2009), in a hedge fund sample, discuss how investable higher moment factors can lead to biased alphas. Many newer models build on one of the two concepts. While the Sharpe ratio and other total risk correction models will require a benchmark in order to evaluate the result, alpha models can be evaluated directly. The Sharpe ratio, for example, is usually interpreted in contrast to the market Sharpe ratio.

Relaxing the Sharpe ratio assumption of quadratic utility will directly lead to a utility based certainty equivalent model as used in Morningstar (2007). Scott and Horvath (1980) show, how higher moment aversion is captured in utility models of higher order than quadratic. As the Sharpe ratio, this utility model also requires some benchmark in order to interpret its outcome. Morningstar (2007) assign a certain investment style to each fund and benchmark funds within the same style box with each other. The DGTW asset by asset style assignment is a lot more accurate. Further, it allows to compare all funds with each other, not only funds within a certain style box. Additionally, it allows for a better analysis of style timing. Goetzmann et al. (2007) show that the Morningstar measure is proof to manipulation.

The Carhart (1997) alpha relaxes the Jensen (1969) assumption of normality by adding three additional factors, size, value and momentum, to the regression. While this model can explain considerable variation in returns, it is not clear if the additional factors really reflect systematic risk[11]. Further, there is still some — possibly undiversifiable idiosyncratic — risk left in the residuals. Among others, Lehmann (1990) and Malkiel and Xu (2002) show that, if

[11] Alternatively, the betas can be interpreted as weights of the factor mimicking portfolios. Then, the Carhart model would reduce to a simple benchmark model

some investors are constrained from holding the market portfolio, they will be forced to care about total risk to some degree in addition to the market risk. They claim that it is what they call *"undiversified" idiosyncratic* risk that explains the cross-sectional difference in equity returns. Goetzmann and Kumar (2008) show that idiosyncratic risk in investor portfolios leads to a welfare loss. The information ratio or Treynor appraisal ratio account for idiosyncratic risk by dividing the Jensen alpha by the residual standard deviation.

Naturally, the question arises whether returns are skilled or lucky. A popular way to test if risk adjusted outperformance is a result of skill and not of luck, is to look at the persistence of these outperformance measures over time. The argument is that luck, contrary to skill, is not persistent. Further, only if the measures are persistent, they can serve as a decision criterion for investors seeking to allocate assets to funds.

For example Hendricks et al. (1993) find evidence on persistence over short-term horizons. Carhart (1992) finds some persistence in mutual fund performance, but attributes it to expense ratios and not to skill. Carhart (1997), evaluating persistence in his four factor alpha, finds that most of the persistence can be explained by the momentum effect[12]. The only unexplained persistence he finds is in the negative alpha part of his sample. Carhart concludes that "persistence in mutual fund performance does not reflect superior stock-picking skill".

Similar papers have been discussing persistence of the Morningstar star ratings [13]. The result by Morey and Gottesman (2006) regarding the predictive power of Morningstar's ratings stands in contrast to the findings of Carhart. Morey and Gottesman (2006) detect successive outperformance of higher rated (and therefore higher-utility) funds, and attribute this effect to the Hendricks et al. (1993) hot hand hypothesis. They do not find support of expense ratios as drivers of that persistence. We know from Carhart that in a factor model, Jegadeesh and Titman (1993) momentum effect can falsely induce the hot-hands-hypothesis. This effect is of course not limited to factor models. Morey and Gottesman (2006) fail to correct for momentum.

The downside of testing for persistence to determine if managers are skilled is twofold. First,

[12]The momentum effect, basically, makes luck persistent. If winners are held by pure chance (and not by trading on the momentum effect or by any skill), and past winners are likely to be future winners, luck will be – to some degree – persistent.

[13]See, e.g. Antypas (2009)

it requires a very long time series to state with reasonable certainty if a manager is skilled. Hence, this measure is usually only suitable to determine whether there are skilled managers in the sample as compared to determining, if a specific manager is skilled. Therefore, it also does not enable to quantify to skilled outperformance in a certain period.

Another way to determine if managers are skilled is the Cremers and Petajisto (2009) active share measure. They measure the activity of a fund manager as deviation of the portfolio's holdings from its closest matching benchmark on a purely descriptive basis, i.e. without directly evaluating the activity. Cremers and Petajisto (2009) find that, the more actively a fund manager deviates from his closest benchmark, the better his outperformance in terms of the Carhart alpha. This direct relation can be viewed as an indication of stock selection skills of active fund managers. Increased activity comes, of course, at the cost of less than perfect diversification. By ignoring these costs, the Carhart alpha used by Cremers and Petajisto might overestimate the benefit of active management.

Fama and French (2010) describe a cross-sectional bootstrapping approach. This approach is primarily suitable to test if, in the entire distribution of fund returns, funds with true positive performance exist. Using the three factor Fama & French alphas and the four factor Carhart alpha, Fama and French (2010) determine if there are more fund managers in the extreme tails of the distribution than pure chance would suggest. They find that true alpha is negative for most if not all funds, yet they cannot rule out that there are a few truly skilled funds. This measure does not, however, qualify as a measure of individual performance, since "good funds are indistinguishable from the lucky bad funds that have negative true alpha".

Barras et al. (2010) use a very similar methodology to determine if there are truly skilled funds in the cross-section. They find that roughly three quarters of all funds are true zero alpha funds and 24% of funds have negative true alpha. Only an insignificant 0.6% of all funds seem to be skilled, this portion increases to 2.4% if short term alphas are considered. In their analysis, fat tails in returns might induce an overestimation of the share of skilled managers. Additionally, as in Fama and French (2010), their analysis does not allow to identify individual skilled funds. At best, it describes a way to identify fund portfolios with a large share of skilled funds.

A detailed examination of manager skill is also delivered by Daniel et al. (1997) as described in section 1. They find stock selection ability in their mutual fund sample. Similarly, Wermers (2000) shows that funds choose stocks that outperformed their characteristic benchmarks by an average of $71bps$ per year on a value weighted basis and $101bps$ per year if equally weighted. Characteristic timing accounts for $2bps$ p.a. Alexander et al. (2007) use the same benchmark portfolios but consider only trades that have opposite sign of the fund flows. This is a more accurate proxy for informed trades than the simple portfolio reallocation proxy by Daniel et al. (1997) and Wermers (2000) which is also used in this paper. On the downside, the measure is likely to miss most supposedly informed trades. Alexander et al. (2007) also find significant skill. All three articles share to common problem of missing risk correction in case the 125 characteristics sorted portfolios do not completely characterize the risk profile. I solve this problem by applying the MPPM to the returns.

Daniel and Titman (1997) and Daniel et al. (2001) discuss whether the sorting is based on factors or characteristics. They find that the returns related to the Fama and French (1993) portfolios cannot be viewed as compensation for factor risks but are based on high covariances within these portfolios. These high covariances reflect the fact that firms within these portfolios tend to have similar properties. As long as the explanatory power in the cross section is large, in contrast to the above mentioned articles, it does not matter to my model whether Fama & French have identified factors in the Merton (1973) sense (explaining consumption growth that is orthogonal to the market) and therefore reflecting systematic risk or just characteristics explaining the cross-section of returns. To correct for characteristic luck, it is sufficient that assets in the same characteristic portfolio strongly covary.

3 Methodology

Goetzmann et al. (2007) define criteria to measure the value added by active management. First, the measure has to be proof to manipulation from an ex-ante point of view — an uninformed manager should not be able to pretend outperformance by loading on priced risks not captured by the measure[14]. Goetzmann et al. (2007) show that, by assuming "nice" return distributions, traditional measures like the Sharpe ratio or the Jensen (1969) alpha can be manipulated. In addition, because active investment might lead to less than perfect diversification, idiosyncratic risk needs to be taken into account. Also, because ex-post returns are evaluated, the measure should be proof to luck — an uninformed manager should not be credited outperformance that was simple luck. After all corrections, the measure should reduce to a single dimension. The score not to depend on the portfolio's dollar value and to be consistent with financial market equilibrium conditions.

3.1 The Utility Based Skill Measure

Goetzmann et al. (2007) show that their MPPM of the form

$$MPPM(r_t) = \frac{1}{(1-\rho)\Delta t} ln(\frac{1}{T}\sum_{t=1}^{T}(\frac{1+r_t}{1+r_{ft}})^{1-\rho}) \quad (1)$$

can match their criteria [15]. The MPPM is the average of a power utility function and — for an investor following this type of utility — gives the average annualized certainty equivalent excess return. I will work with monthly observations ($\Delta t = 1/12$).

Expressing risk adjusted returns in terms of their certainty equivalent is intuitively much more understandable to the investor than comparable measures, such as the Sharpe ratio. Goetzmann et al. (2007) find that their MPPM is "identical in substance and nearly in form to the Morningstar Risk Adjusted Rating (MRAR)" methodology. The measure requires a suitable risk aversion coefficient. Goetzmann et al. (2007) suggest to calibrate it in the benchmark index such that it is ideal for an uninformed investor to hold that benchmark. Therefore, ρ is defined as

$$\rho = \frac{ln[E(1+\tilde{r}_b)] - ln(1+r_f)}{Var[ln(1+\tilde{r}_b)]} \quad (2)$$

[14] It needs to be distinguished between ex-ante uninformed score enhancement — manipulation, and ex post uninformed score enhancement — luck. I correct for both.

[15] For a more complete discussion of the criteria a performance measure of active management has to match, see Goetzmann et al. (2007).

In my sample, with data from 1983 to 2010 and the CRSP value weighted index as the benchmark, I obtain $\rho = 2.7$. This risk aversion coefficient will be used for all further calculations. It is slightly below the coefficient used by Morningstar[16]. Since only ex-post data can be analyzed, after correcting for risk, those returns that can be attributed to the fund manager's active investment decisions have to be separated from those that were pure luck. Without knowing the fund manager's motivation to hold an asset, distinguishing lucky and skilled returns is difficult, or as Fama and French (2010) put it, "unfortunately, (...) good funds are indistinguishable from the lucky bad funds that land in the top percentiles (...) but have negative true alpha". Borrowing from Daniel et al. (1997), investment decisions are proxied for with trades. Timing and stock picking performance are expressed as differences in certainty equivalent. The manipulation proof timing measure therefore is

$$MP_{timing} = MPPM(r_t^{\tilde{b_{t-1}}}) - MPPM(r_t^{\tilde{b_{t-13}}}) \tag{3}$$

with

$$r_t^{\tilde{b_{t-1}}} = \sum_{j=1}^{N} w_{j,\tilde{t}-1} R_t^{\tilde{b_{j,t-1}}}$$

and

$$r_t^{\tilde{b_{t-13}}} = \sum_{j=1}^{N} w_{j,\tilde{t}-13} R_t^{\tilde{b_{j,t-13}}}$$

and MP-selectivity

$$MP_{selection} = MPPM(r_t(holdings)) - MPPM(r_t^{\tilde{b_{t-1}}}) \tag{4}$$

with

$$r_t(holdings) = \sum_{j=1}^{N} w_{j,\tilde{t}-1} \tilde{R_{j,t}}$$

and

$$r_t^{\tilde{b_{t-1}}} = \sum_{j=1}^{N} w_{j,\tilde{t}-1} R_t^{\tilde{b_{j,t-1}}}$$

[16]If ρ is set to 3, the MPPM reduces to the Morningstar risk adjustment methodology.

I use the 125 size-, value- and momentum sorted portfolios as suggested by Daniel et al. (1997) to create an individual, style adjusted benchmark for every fund. The more uniformly assets within these benchmark portfolios react to lucky events, the better suitable are these measures as luck filters. Fama and French (1993) show that the prices of high book-to-market and small size stocks move up and down together.

3.2 Assumptions

As many of the assumptions of prior performance measures could lead to manipulation of their results, the new measure requires only rather lax assumptions:

1. No assumption is needed on types of priced risk as all risks are treated the same.
    (a) Systematic risk can be single or multi-dimensional, i.e. higher order risk can be priced or not.
    (b) People can be averse to idiosyncratic risk if they cannot hold the market portfolio and at the same time maintain active. Managers, who assemble an alpha generating portfolio that has to much excess risk, can be skilled in selecting mispriced securities without generating any value for the investor.
2. We have to assume that luck is correlated in the sense that similar stocks perform similarly in case of unpredictable events. Put differently, the benchmark based luck correction is limited to detecting characteristic luck.
3. Return distributions are not parametrized, therefore no assumptions regarding the distribution of returns are required. However, a small sample of only 12 annual observations can lead to misestimations of the true moment exposure, especially of degree higher than two as realizations in the tails are unlikely to occur (but can have severe economic consequences).
4. Power utility sufficiently characterizes the investor's utility. However, parameters can be easily modified in the model maintaining some flexibility in the utility assumption. Especially, results can vary in the sense that managers do create some excess utility for investors not so risk averse but do not for very risk averse investors.

3.3 Limitations

Above assumptions lead to some limitations of the methodology. Specifically, by proxying for informed decisions only with reallocations, some luck might be attributed to skill. Further, the measure could potentially miss out on risk.

As a result of the brevity of the sample used to calculate the annual certainty equivalent excess return ($T = 12$ in equation 1), the true moment exposure of the portfolio could remain opaque. Especially higher moment risk could require a significantly longer sample in order to be captured by the utility function. A similar moment exposure of all assets in the same benchmark portfolio would mitigate this problem, but following Daniel and Titman (1997), we cannot assume a uniform moment exposure of the benchmark portfolios. Further, managers could deliberately load on higher moments, hoping the exposure would stay undiscovered. With regard to the identification of the risk-costs of stock picking, this effect works in favor of my results. If risk is not fully captured, the costs are — at most — underestimated[17].

The MPPM assumes the investor follows power utility with $\rho = 2.7$. This is a rather basic utility model that ignores the properties of Kahneman and Tversky (1979) prospect theory. Notably, the investor's aversion to negative returns might be underestimated, consequently overestimating the certainty equivalent excess return. Again, this effect only works in favor of the discovery of risk-cost of stock picking.

Further, regarding the timing measure, reallocation as a proxy for informed trades could be imprecise in some instances. Since the performance of this year's benchmark is compared to the performance of last year's benchmark, the methodology will not credit informed passivity of the portfolio as a skill to the manager but will falsely attribute it to luck. For example, a manager with information about a longer term outperformance of a certain set of characteristic portfolios will only be credited the resulting performance in the first period after the reallocation. In all succeeding periods, the possible outperformance will be considered luck, thus underestimating the manager's skill[18]. In other words, if the manager actively decides to stay on the same

[17] In choosing the evaluation interval, a balance between achieving an adequate sample length and minimizing the survivorship bias has to be found. Choosing $T = 12$ in equation 1 allows the sample to reset once a year.

[18] Multi-period or persistent outperformance can be attributed to the momentum effect. As this is a known

benchmark and the old benchmark performs well, this performance will not be considered timing skill. "Active passivity" is not rewarded by the DGTW timing measure.

Finally, portfolio reallocations could be considered informed if they, in truth, were not. For example, a manager could falsely identify reasons for outperformance of certain characteristics and trade into these characteristics. However, these reasons could prove wrong and some other, unpredicted occurrence could lead to superior performance of the held characteristics. In this case, luck will be falsely attributed to skill. Therefore, the measure introduced in this paper does still not allow to state with absolute certainty if an individual manager has skill or not. This has no implications for the results regarding the risk-costs of stock picking. Alexander et al. (2007) propose a more accurate proxy of informed trades, considering only those reallocations with opposite sign of the total fund flow. While trades identified by this methodology are informed with relative certainty, all informed trades that have the same sign as the total fund flow are not considered. Thus, this methodology might be suitable to answer the binary question if a manager is, at all, skilled. To quantify the skill, however, we need to include those trades as well, even at the risk of including some luck into the skill measure.

3.4 Multi Fund Portfolios

To analyze the diversifiability of the risks introduced by stock picking, multi fund portfolios are assembled in the following fashion: 500 hypothetical investors randomly draw at the end of each year from all funds that will survive the entire following year[19] (without replacement, as investors are unlikely to buy the same fund twice). Investors will equally weight[20] all funds they draw into their portfolio and hold them for the entire following year. Afterwards, investors draw again (from the sample now containing all funds that survive the successive year). I simulate nine different portfolio sizes with 1, 2, 3, 5, 8, 10, 15, 20 and 30 funds. This results in 4500 different portfolios to analyze every year. To ensure comparability, one fund portfolios are assembled

effect, it is arguable if this should be considered skilled outperformance that justifies the high fees of active management. This mitigates the mentioned limitation.

[19]While this introduces a survivorship bias, we need this assumption to maintain consistency with the simple one fund measure. In the simple one fund evaluation, certainty equivalents can only be computed for funds that survive the entire year (or that have data available the entire year).

[20]Alternative specification: value weight.

with the same methodology[21]. Results are averaged over all 500 randomly selected portfolios.

3.5 Herding

If herding explains the undiversifiable part of the selection spread, I would expect the portfolio selection spread to decrease over time as prior research by Sias (2004) has shown that herding was more severe in the 1980s than in the 1990s. I find this pattern in the data. For a more thorough analysis of the influence of herding, I construct a new herding measure, which focuses on herding in holdings, not in trades. The prior standard developed by Lakonishok et al. (1992) analyzes only trades. To measure the impact of herding on portfolio diversifiability, I have to look at all holdings including those that were static during the past period.

The herding measure is constructed as a direct derivation of the active share measure by Cremers and Petajisto (2009), only that it measures the activity of the entire identifiable fund universe compared to the CRSP value weighted index, instead of the activity on a single fund level. If this derivation is large it means that funds as a whole overweight a specific asset or a specific group of assets severely compared to the market weight of the asset. Consequently, the herding measure is

$$Herding = \frac{1}{2}\sum_{i=1}^{N} |\omega_{funduniverse,i} - \omega_{market,i}| \tag{5}$$

with $\omega_{funduniverse,i}$ the percentage of all known fund holdings invested in asset i and $\omega_{market,i}$ the weight of asset i as a percentage of the total market. I calculate the herding measure every month and use the maximum herding measure in every year to explain the selection spread, which is annual data.

If herding is extreme, investors cannot construct a well diversified portfolio simply from mutual fund investments.

[21]The results from the simulated one fund portfolios slightly differ from the regular fund by fund analysis. This small divergence is pure chance and based on the outcome of the random draws.

4 Data

Monthly return data and fees are obtained from the CRSP survivor bias free mutual fund database. Equity holdings portfolios are obtained from Thomson Financial's CDA/Spectrum database. Stock and market index return data are from CRSP. Fama & French and momentum returns are from Kenneth French's data library[22]. The 125 characteristic sorted portfolios are available on Russ Wermers' website[23]. For a detailed description of the databases, see e.g. Carhart (1997), Daniel et al. (1997), or Wermers (2000).

I include only funds with investment objective codes 2, 3, 4 or 7[24]. Further, to be included in the sample, at least 80% of the fund's reported TNA has to be identifiable in terms of CDA/Spectrum holdings. Holdings reports are quarterly, sometimes semi-annually. It is assumed holdings stay constant after each report date until the next report. If gaps between reports are larger than six month, the fund is not included in this time period. Returns in the holdings portfolio are windsorized at 1% level based on their spread to the reported gross returns of the fund.

Funds are only included in years with full data availability because the calculation of the certainty equivalent measure requires a full year of data. Since the focus of this paper is on the quantification of risk-costs introduced by stock picking and market timing, and not on the absolute size of stock picking or timing skills, funds that have never exceeded USD five million in assets and are therefore possibly exposed to an incubation bias[25] are not excluded from the sample for the sake of sample size. Stock picking or timing risks of funds should be comparable whether generated prior to or after their incubation.

To calculate the timing measure, data needs to be available for two consecutive years, as a ramp up period of one year is required. In order not to sacrifice to much data, the same condition is not imposed on the selection measure. The first data point on timing is ergo only in 1984 instead of 1983. Hence, the timing and the selection measures are not calculated on the

[22] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[23] http://www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm
[24] Investment objectives aggressive growth, growth, growth & income or balanced.
[25] See, e.g. Fama and French (2010).

same data base. Again, since I do not draw conclusions about the absolute size of timing and selection skills, nor compare timing and selection skills, this should not be an issue.

CRSP monthly returns are on share class level. To analyze portfolio returns, I aggregate the share class returns to portfolio level.

4.1 Summary Statistics

Summary statistics are reported in table 1. The average annual selection measure is 87*bps* per year, slightly below the equally weighted 101*bps* reported by Wermers (2000) in a different sample. The difference could be due to the advend of index funds or more efficient financial markets. The difference between net fund returns and returns to the holdings portfolio is 202*bps*, also slightly smaller than the 280*bps* reported in Wermers (2000). 121*bps* of this difference can be attributed to fees, the rest is likely do to expenses and lower returns on the non-stock portfolio. Keep in mind that this difference was windsorized. Index returns are, on average, 45*bps* higher than the net fund returns, but 115*bps* lower than the returns of to the fund's holdings portfolio, compared to 130*bps* reported in Wermers (2000). Average annualized timing was 8*bps*, about the same as before. Keep in mind the numbers in table 1 overweight small funds. I have a total of 21,686 funds in my sample, except in the timing sample, were only 15,782 funds remain.

The herding measure does not show much variation in the time series, it is on average 33%, meaning that fund managers to some degree move into the same direction and, on aggregate — deviate from the market weighting of funds. In line with the results by Sias (2004), who use an entirely different measure, herding seems to decrease over time, most likely with the advent of index funds that can be seen in table 4, but peaked again in 2009 after the financial crisis. The standard deviation of the market over my sample period is 15.83% on average, in single years it is as low as 7.42% and as high as 40.28%. Returns seem to be slightly negatively skewed and have fat tails. Interestingly, kurtosis can be as little as 2.76 in 2005 and as high as 41 in 1987, the year of the Black Monday CPPI crash. Other extreme kurtosis years are in 1988, 1989 and 1997.

Table 1: Equal weighted summary statistics (%).

The four factor alpha is obtained by a regression of the 12 monthly returns each year on the Fama & French and momentum factors. DGTW Time and DGTW Sel are the timing and selection measures from Daniel et al. (1997). Note that alphas are monthly and all other returns are annualized. Alphas are based on net returns, the DGTW selection measure is based on holdings returns and therefore before fees and trading costs. The last line contains time series averages of the measures, except for the number of funds, where it reports the total number of fund years. Ret is the annual return of the CRSP value weighted index. Sd, Skew and Kurt are standard deviation (%), skewness and kurtosis of the daily returns of the CRSP value weighted index. Daily returns are used to capture as much of the volatility of the index as possible and the be able to determine skewness and kurtosis with some reliability.

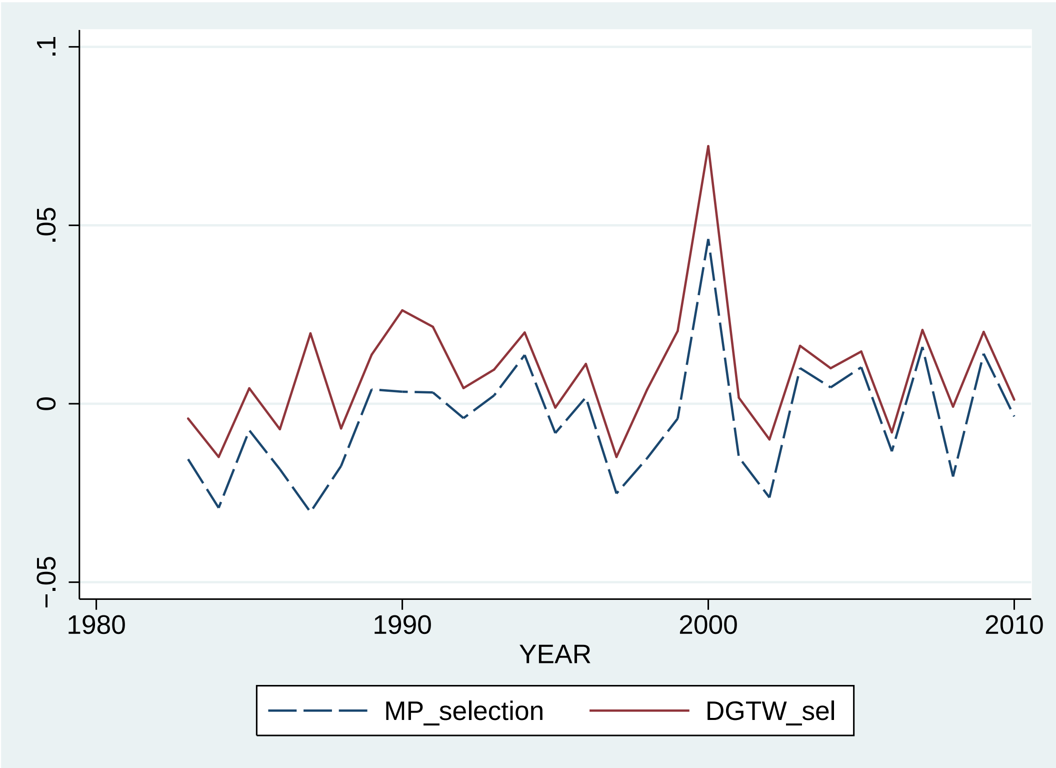
	Funds												
	No	Fund Ret	Holdings Ret	DGTW Sel	Alpha	Fees	No Time	DGTW Time	Herding	Ret	Sd	Skew	Kurt
1983	106	19.83	22.43	-0.41	0.08	1.03			41.03	22.64	12.11	-0.15	3.53
1984	127	-2.76	-0.32	-1.49	-0.04	0.97	73	-1.50	38.66	3.04	11.35	0.88	4.63
1985	140	28.63	31.93	0.43	-0.29	0.99	88	0.17	37.13	31.38	8.88	0.34	3.27
1986	186	13.21	13.49	-0.71	0.05	1.01	115	-0.42	34.70	15.65	12.60	-1.08	7.03
1987	175	0.02	1.86	1.97	-0.08	1.02	129	0.21	38.34	1.74	27.89	-3.74	41.88
1988	230	15.03	17.53	-0.69	0.08	1.22	137	-0.46	35.08	17.61	14.01	-0.96	10.60
1989	245	26.73	29.85	1.38	0.19	1.29	177	0.44	34.29	28.48	10.88	-2.11	18.10
1990	222	-6.38	-6.05	2.62	0.31	1.26	162	0.76	35.48	-6.08	14.35	-0.28	4.18
1991	269	38.75	42.80	2.16	-0.16	1.15	176	-0.90	34.37	33.78	12.88	0.12	5.16
1992	329	9.42	11.15	0.44	-0.07	1.27	205	-1.02	32.22	9.07	8.98	-0.05	3.44
1993	343	12.24	13.49	0.96	0.01	1.18	220	0.28	33.28	11.60	7.97	-0.67	6.42
1994	405	-1.15	1.90	2.00	-0.04	1.16	234	0.47	34.88	-0.64	9.21	-0.39	5.18
1995	609	32.32	36.01	-0.11	-0.17	1.22	319	0.62	35.38	35.74	7.42	-0.39	4.33
1996	630	19.25	22.60	1.11	-0.04	1.24	388	0.80	36.82	21.23	10.86	-0.67	5.00
1997	922	24.85	27.24	-1.49	-0.29	1.24	513	-0.29	33.90	30.43	15.82	-0.98	10.75
1998	896	16.72	18.78	0.39	-0.15	1.29	616	0.57	33.48	22.34	19.48	-0.62	7.53
1999	1,065	29.15	27.57	2.04	0.13	1.29	674	4.80	33.10	25.66	17.05	-0.01	2.93
2000	1,163	1.24	4.97	7.22	0.79	1.29	812	-0.34	31.74	-11.22	24.53	0.04	4.17
2001	1,182	-8.68	-6.24	0.17	-0.14	1.33	828	-1.15	28.09	-11.06	21.82	0.08	4.59
2002	1,326	-22.33	-20.34	-1.00	-0.32	1.39	890	-1.27	26.08	-20.89	24.55	0.47	3.68
2003	1,681	35.08	37.48	1.62	-0.43	1.38	1,150	-0.24	26.07	33.15	15.99	0.06	3.58
2004	1,632	12.61	14.62	1.00	0.03	1.34	1,350	0.00	25.58	13.02	11.23	-0.21	2.88
2005	1,612	7.44	9.21	1.46	0.02	1.27	1,347	-0.98	25.89	7.32	10.30	-0.06	2.76
2006	1,647	12.91	13.48	-0.80	-0.04	1.25	1,381	-0.12	27.20	16.22	10.61	0.17	4.19
2007	1,446	7.66	8.87	2.07	0.11	1.20	1,333	1.14	27.61	7.38	15.83	-0.45	4.18
2008	1,008	-38.40	-36.97	-0.08	0.01	1.21	877	1.97	28.89	-38.22	40.28	0.09	6.65
2009	1,111	33.86	34.59	2.01	0.14	1.21	778	-1.43	40.61	31.60	28.02	-0.03	4.62
2010	979	19.10	21.01	0.11	-0.19	1.14	810	0.05	34.43	18.03	18.48	-0.19	4.97
Mean/Total	21,686	12.01	14.03	0.87	-0.02	1.21	15,782	0.08	33.01	12.46	15.83	-0.39	6.79

5 Single Fund Portfolio Results

In this section, I will first quantify the risk cost of stock picking (section 5.1 and than try to explain what drives this risk in the cross section (section 5.2), in the time series (section 5.3).

5.1 Risks incured by stock picking

Figure 1: Equal weighted DGTW and MP selection measures 1983–2010.



In figure 1, the risk adjusted certainty equivalent based selection measure (MP selection, henceforth) is graphed against the traditional DGTW selection measure. As hypothesized, the stock picking skill as measured by the DGTW selection measure is significantly (economically and statistically) larger. Further, there is a large variation of the average spread in the time series. The equal weighted selection spread is also reported in table 2. The average equal weighted selection spread is 131*bps* from 1983 to 2010, and is statistically different from zero ($p < 0.0001$). This spread peaks in 1987, with other high realizations in 1990 and 1991 and then again from 1999–2002. In the the early 1990s and 2000s, the spread decreases to below one percent, before it peaks again in 2008. In a value weighted sample, the spread is only 94*bps* on average. As discussed in section 5.2, the difference to the equally weighted spread is entirely due to lower idiosyncratic risks taken by high TNA funds. From table 2, it becomes further apparent that there is extremely large cross sectional variation in the spread, with the difference between the maximum and the minimum spread decile almost 9% on average. This cross sectional variation is highly persistent, with the difference of the top and bottom decile portfolios sorted by the lagged spread being 494*bps* after one year and still 320*bps* per annum after five years.

Taking a closer look at the persistence table (2) suggests that the spread is a persistent characteristic of the individual fund and only slowly changes over time. In other words, if the spread is a measure of unidentified risk taking, some funds consistently expose themselves to these unidentified risks, while others are hesitant to do so. As a result, investors in some funds are persistently and unknowingly exposed to higher risks than others. This makes it all the more important to quantify these risks. A look at the long term persistence in the table also allows a first inference about the time series variation of the spread. Consider, for example, the realizations in the 1983 line. While the difference between the top and bottom portfolio is comparably low 4.69% in the first year and around 2% in the following years, in $t = 4$ this difference increases to 8.79% in performance per annum. $t = 4$ refers, in this case, to 1987, the year with the second highest performance gap in the sample. Apparently, we can already see in 1983 if a fund will perform particularly bad in terms of its unidentified risk exposure four years later, in 1987. The same pattern holds for $t = 3$ in 1984 and so on. My hypothesis is that the particularly strong punishment of risk taking in some years, such as 1987, is likely to be explained not only by changing risk exposure of the individual funds, but also by the way these higher moment risks can be captured by the MPPM. Recall from section 3.3 that, using only twelve monthly data points to compute the certainty equivalents, I might miss a lot of higher order risk exposure in some years, because there are just no extreme realizations in the data. Contrarily, there might be years with an above average number of extreme realizations, leading to an overestimation of risk exposure and thus selection spread. The time series average over 28

Table 2: Difference between the top and the bottom decile selection spread each year and persistence of the difference (in %).

The first two columns give the average, equal weighted, DGTW- and MP selection measures in each year. Column 3 gives the difference between the decile of funds with the highest selection spread and the decile of funds with the lowest selection spread. Columns 4 to 9 check the persistence of this difference. Therefore, e.g. in column 8 (t=4) in year 1983, the difference between the decile of funds that had the highest and the lowest selection spread in 1983 is given for 1987. The last column reports the value weighted results in $t = 0$

	MP Sel	DGTW Sel	Sel Spread	Decile 10 - Decile 1						VW
	t=0	t=0	t=0	t=0	t=1	t=2	t=3	t=4	t=5	t=0
1983	-1.55	-0.41	1.14	4.69	2.33	2.02	2.02	8.79	1.51	0.90
1984	-2.92	-1.49	1.43	6.33	3.82	3.50	12.65	1.25	3.32	1.13
1985	-0.74	0.43	1.18	5.21	2.57	13.28	0.92	2.26	3.40	1.01
1986	-1.83	-0.71	1.12	4.60	5.69	0.23	2.67	6.15	4.81	0.83
1987	-3.04	1.97	5.01	21.18	1.83	1.67	4.31	3.05	0.48	4.75
1988	-1.75	-0.69	1.05	3.72	1.34	6.06	3.32	2.43	1.82	0.79
1989	0.40	1.38	0.97	4.01	4.99	4.97	1.52	0.27	0.76	0.72
1990	0.34	2.62	2.28	10.38	4.76	3.36	1.48	1.86	1.79	1.53
1991	0.32	2.16	1.84	7.77	2.88	2.33	3.25	1.84	3.63	1.43
1992	-0.40	0.44	0.84	3.87	1.97	3.70	2.83	4.46	3.93	0.68
1993	0.23	0.96	0.72	2.94	3.08	2.54	4.35	2.42	6.00	0.54
1994	1.37	2.00	0.63	3.69	1.93	2.65	1.91	5.29	4.37	0.58
1995	-0.82	-0.11	0.71	3.51	3.21	3.32	6.55	4.27	5.96	0.64
1996	0.18	1.11	0.93	5.21	3.99	7.47	5.13	8.97	7.95	0.79
1997	-2.51	-1.49	1.02	6.09	9.19	6.27	8.77	10.84	5.42	0.76
1998	-1.53	0.39	1.92	14.22	7.07	10.60	7.36	3.33	0.39	1.29
1999	-0.42	2.04	2.45	13.91	11.38	9.45	6.07	1.64	1.82	1.51
2000	4.62	7.22	2.60	22.97	11.49	5.75	1.99	1.70	1.32	1.84
2001	-1.50	0.17	1.67	20.36	9.87	2.59	1.67	1.18	0.96	1.02
2002	-2.63	-1.00	1.63	13.42	3.15	2.00	1.60	1.35	0.73	0.75
2003	1.00	1.62	0.63	4.93	2.80	2.01	1.69	1.01	6.28	0.20
2004	0.46	1.00	0.54	3.86	2.34	2.20	1.37	8.01	1.29	0.26
2005	1.02	1.46	0.44	3.08	2.00	1.36	7.29	2.97	1.50	0.20
2006	-1.33	-0.80	0.53	3.05	1.49	7.30	0.49	1.68		0.29
2007	1.59	2.07	0.48	3.10	8.73	3.76	1.96			0.22
2008	-2.04	-0.08	1.96	14.08	4.77	2.27				1.15
2009	1.40	2.01	0.61	10.09	2.11					0.19
2010	-0.36	0.11	0.47	4.50						0.31
Mean	-0.44	0.87	1.31	8.03	4.47	4.33	3.73	3.63	3.02	0.94

Table 3: Summary Statistics sorted by the lagged Selection Spread decile (%).

Funds are sorted after their past selection spread top and bottom deciles are further divided into terciles. Market beta is the coefficient of the market factor of the returns of the fund's holdings portfolio on the market, size, value, and momentum factors as suggested by Carhart (1997). RMSE is the root mean squared error of a the same regression.

	MP Sel	DGTW Sel	Market Beta	RMSE	Sel Spread
1A	1.30	0.06	0.79	1.59	-1.24
1B	0.72	0.65	0.95	1.35	-0.06
1C	-0.42	1.07	1.05	1.87	1.49
1	0.54	0.59	0.93	1.60	0.05
2	0.22	0.41	0.97	1.23	0.20
3	0.33	0.60	0.98	1.08	0.27
4	0.44	0.79	0.99	1.21	0.35
5	0.39	0.84	0.99	1.31	0.45
6	0.47	1.06	1.00	1.42	0.59
7	0.56	1.38	1.03	1.54	0.81
8	0.33	1.51	1.04	1.76	1.18
9	-0.88	1.07	1.09	2.19	1.96
10	-2.69	2.11	1.21	3.09	4.80
10A	-0.26	0.59	1.04	2.32	0.86
10B	-3.04	0.71	1.21	2.94	3.75
10C	-4.87	5.11	1.39	4.03	9.98
10C-1A	-6.17	5.06	0.59	2.44	11.22
10-1	-3.23	1.52	0.28	1.48	4.75

years should roughly balance. In section 5.3, I find a way to proxy for the amount of risk that can be captured by the data, and analyze if the time series variation is indeed explained by this capturability.

In table 3, key characteristics of funds, sorted by their lagged selection spread, are presented. Funds with a large selection spread in the past year have higher DGTW selection measure and lower MP selection measure. They seem to "buy" the falsely attributed stock pick skill by taking higher risks, but they seem to overpay. The additional risk only leads to 1.52% higher performance (top minus bottom decile DGTW), but risk costs are 4.75% higher (top minus bottom decile sel spread) compared to lowest decile funds that do not take these risks. It can also be seen that managers buy higher stock picking skills as measured by the DGTW selection measure buy taking higher market risks (beta is 1.21 in the top decile compared to 0.93 in the bottom decile) and that managers take a lot more unobserved risks as measured by the residual error in the Carhart 4-factor regression. To evaluate if these unobserved risks are diversifiable, the multi-fund portfolio analysis will be conducted. But first, in the following section, a look at the cross section of funds has to explain, what high-stock-picking-risk funds do differently from low-risk funds (apart from taking more beta risk).

The large economic consequences of this risk exposure suggest a more thorough analysis is worthwhile. Particularly, as mentioned, I will analyse if these risks can be avoided by holding well diversified fund portfolios, maybe because funds in the top spread deciles are extremely active and bet heavily on few single stocks or if these funds expose themselves to some undiversifiable risk.

5.2 Analysis of selection risks in the cross section

In this section, I take a closer look at the cross section. I analyze the different risk loadings by fund investment style. In a second step, by building the multi-fund portfolios described in section 3.4, I will quantify the undiversifiable part of the selection spread. I analyze if this undiversifiable part is due to known or unobserved systematic risk taking. I will leave the identification of these risks to further research.

The selection spread varies between different investment styles. It is particularly high for funds focusing on tech firms. This number is, however, driven by the late 1990s and early 2000s, where tech stocks experienced a boom and later a bust. It is noteworthy that the risk was detected already in 1998 and 1999 with a certainty equivalent selection profit 10% and almost 20% below the risk unadjusted selection profit. Note the small number of funds in the sample in these years. Further, it is possible that tech funds are particularly bad diversified. This would drive up the spread but not impose a real cost to the investor, as, by holding a broader fund portfolio, he could avoid that risk. The selection spread of funds focusing on financial firms is also above average. Again, this result is driven by the spread in crisis years 2008 and 2009. Again, sample size is very small. Possibly, higher order risk exposures are just captured particularly well by the MPPM in those years. Not surprisingly, index funds have a below average selection spread, which is close to zero or even negative in most years. As index funds should neither take idiosyncratic risks nor systematic risks in excess of their benchmark, this result confirms the validity of the risk correction measure. The small spread in most years can be explained by the characteristic benchmarks being slightly different from the benchmarks replicated by the fund. Therefore, some variation around zero is inevitable. Small Cap funds have a fairly constant selection spread of around 1% per year. Contrary to all other investment styles, the selection spread of Small Cap funds does not peak in 1987. In light of the hypothesis that time series variation of the selection spread is due to the capturability of the higher order risks in the data, this suggests that small cap funds are the only investment, where managers did not load on the types of higher order risks that became apparent in 1987, or the DGTW benchmarks sufficiently correct for those risks. The remaining spread would than be likely due to idiosyncratic risks in the funds. Large Cap funds deliver results comparable to index funds, and Value and Growth funds behave in a fashion similar to the entire sample average.

Table 4: The average selection spread in each year by type of fund (in %).

The funds are assigned to each group according to their names, i.e. if a fund name contains the word "index" it is considered an index fund. The table reports the average selection spread, sorted by fund type, each year and the number of funds with data available for each type. The last line reports time series averages for the means and time series totals for the fund numbers.

	All		Value		Growth		Small		Large		Tech		Financials		Index		Others	
	No	Spread	No	Spread	No	Spread	No	Spread	No	Spread	No	Spread	No	Spread	No	Spread	No	Spread
1983	106	1.14	2	0.84	28	1.19	1	0.08			2	2.24			1	0.12	72	1.13
1984	127	1.43	4	1.18	30	1.36	1	0.79			1	1.84			1	0.09	90	1.48
1985	140	1.18	4	1.12	35	0.99	1	0.07			4	3.22	2	0.74	1	0.08	93	1.19
1986	186	1.12	7	1.77	40	0.92	2	-0.03			4	1.26	6	1.90	1	0.30	126	1.13
1987	175	5.01	7	3.49	34	5.92	2	-0.16			7	10.18	5	3.53	2	1.37	118	4.74
1988	230	1.05	10	0.97	47	1.25	2	-0.03			6	2.82	8	0.77	4	0.65	153	0.97
1989	245	0.97	10	0.63	54	0.93	3	0.76			3	2.34	4	1.52	5	0.09	166	1.00
1990	222	2.28	9	1.58	56	2.58	5	3.25			4	7.00	3	-1.42	7	0.78	138	2.19
1991	269	1.84	15	1.06	67	1.81	7	2.03			6	6.99	3	5.00	8	0.71	163	1.72
1992	329	0.84	23	0.69	69	0.78	14	1.19	1	0.02	5	3.55	3	2.12	8	0.25	206	0.80
1993	343	0.72	25	0.40	101	0.86	13	1.22	1	0.19	5	1.96	1	-0.39	11	0.17	186	0.67
1994	405	0.63	20	0.30	114	0.65	36	0.82	3	-0.13	5	3.08	1	0.12	20	0.06	206	0.63
1995	609	0.71	41	0.31	164	0.75	65	1.27	13	0.12	2	1.95	1	-0.27	30	0.07	293	0.71
1996	630	0.93	46	0.20	159	1.14	76	1.41	13	0.34	3	4.45	2	0.81	24	0.16	307	0.87
1997	922	1.02	67	0.65	215	1.07	153	1.18	36	0.57	7	6.18	5	2.73	36	0.39	403	0.98
1998	896	1.92	63	0.44	195	2.24	134	2.45	48	0.68	15	11.19	9	5.17	40	-0.03	392	1.74
1999	1,065	2.45	81	1.55	235	2.20	203	2.65	49	1.95	14	18.47	7	3.14	51	0.99	425	2.37
2000	1,163	2.60	91	1.76	258	2.30	229	2.26	50	1.54	16	18.45	9	8.99	60	-0.08	450	2.91
2001	1,182	1.67	94	-0.12	214	1.45	237	1.50	76	0.14	31	25.86	15	0.47	70	-0.44	445	1.19
2002	1,326	1.63	104	0.96	224	1.23	288	1.52	111	0.64	44	16.26	20	0.49	84	0.35	451	1.15
2003	1,681	0.63	134	0.68	284	0.41	391	0.52	142	0.15	53	3.75	23	0.88	106	0.02	548	0.73
2004	1,632	0.54	127	0.22	267	0.56	388	0.37	160	0.20	47	2.82	19	0.37	89	0.02	535	0.72
2005	1,612	0.44	131	0.19	251	0.44	404	0.32	160	0.20	33	1.55	21	-0.01	106	0.12	506	0.70
2006	1,647	0.53	138	0.22	238	0.53	421	0.46	156	0.29	30	1.77	20	0.40	107	0.15	537	0.75
2007	1,446	0.48	127	0.34	198	0.33	361	0.42	144	0.12	30	0.91	12	1.67	97	0.08	477	0.74
2008	1,008	1.96	81	1.45	125	1.68	275	1.60	114	1.14	17	5.02	7	4.19	73	0.17	316	3.00
2009	1,111	0.61	91	0.99	147	0.06	303	0.38	111	-0.12	19	1.49	8	4.75	65	-0.03	367	1.13
2010	979	0.47	69	0.51	133	0.66	263	0.19	95	0.49	18	1.25	9	1.79	71	0.06	321	0.62
Mean/Total	21,686	1.31	1,621	0.87	3,982	1.30	4,278	1.02	1,483	0.45	431	5.99	223	1.90	1,178	0.24	8,490	1.36

Results reported in table 5 show that about two thirds of the annual selection risk is diversifiable, 87*bps* on average. It can be seen that diversification gain is large between a one fund and a five fund portfolio, and is only 1*bps* on average as portfolio size increases from 20 to 30 funds. Depending on their transaction costs, fund investors should hold at least five to eight different funds in their portfolio. The time series mean of the undiversifiable part of the spread is 47*bps*. This brings the average systematic risk adjusted returns to stock picking down to 42*bps* in my slightly survivor biased sample. Even the undiversifiable part of the spread has its peak in 1987, where the average randomly selected 30 fund portfolio suffered risk costs of 3.66%. In years 2001 and 2009, to the contrary, the spread was entirely diversifiable and systematic risk taken by the average fund was below its benchmark. Thus, the traditional DGTW measure underestimated the systematic risk adjusted returns to stock picking in these years. In table 11 I conduct the same analysis for the value weighted sample. The diversification gain in the value weighted portfolios is smaller than in equal weighted portfolios, indicating that higher TNA funds are better diversified. Systematic risk taking in the value weighted sample, measured as the average selection spread of the diversified value weighted portfolio, is slightly larger and statistically equivalent ($p < 0.0001$) to the undiversifiable stock picking costs in the equal weighted sample. I base all my analysis on equal weighted results, because I believe investors are more likely to hold equal- than value weighted portfolios. The results for the systematic risk component of the value weighted stock-picking costs show that this approach dies not influence the results.

Despite the apparent partial diversifiability of the selection spread in table 5, the large cross sectional variation of the spread can only partially be explained by idiosyncratic risk. Cross sectional regression results of a regression of the selection spread on the number of holdings in the fund portfolio and the fund TNA, presented in table 13 in the appendix, do not explain an economically significant portion of the spread with coefficients and R^2 at or close to zero for the all fund sample. This result, however, is not robust across investment styles. For Small Cap, Tech and Financials funds, the number of holdings and the total net assets of the fund do explain some of the cross sectional variation of the selection spread, even though the coefficients are not significant in most years. Considering the large cross sectional variation of the number

Table 5: The average selection spread of multi fund portfolios in each year (in %).
Multi fund portfolios are build as described in section 3.4. MP selection measures are based on the portfolio certainty equivalent. The last column gives the diversification gain as the difference between the selection spread of the 30 fund portfolio and the 1 fund portfolio.

	No. of Funds in Portfolio									**Divers. Gain**
	1	2	3	5	8	10	15	20	30	30-1
1983	1.14	0.63	0.63	0.54	0.48	0.47	0.46	0.43	0.43	0.71
1984	1.57	1.02	1.02	0.87	0.84	0.82	0.83	0.79	0.79	0.78
1985	1.21	0.78	0.78	0.73	0.72	0.73	0.69	0.68	0.67	0.54
1986	1.14	0.74	0.74	0.66	0.63	0.60	0.56	0.58	0.56	0.58
1987	4.99	4.54	4.54	3.95	3.94	3.86	3.76	3.75	3.66	1.32
1988	1.10	0.58	0.58	0.46	0.43	0.43	0.40	0.36	0.37	0.73
1989	0.96	0.56	0.56	0.48	0.45	0.44	0.41	0.41	0.40	0.56
1990	2.23	1.60	1.60	1.51	1.47	1.42	1.43	1.34	1.36	0.88
1991	1.90	1.27	1.27	1.19	1.16	1.18	1.09	1.09	1.11	0.79
1992	0.81	0.46	0.46	0.37	0.33	0.31	0.27	0.27	0.25	0.55
1993	0.76	0.35	0.35	0.27	0.22	0.22	0.20	0.18	0.18	0.58
1994	0.71	0.30	0.30	0.23	0.21	0.21	0.19	0.18	0.17	0.55
1995	0.72	0.33	0.33	0.26	0.22	0.19	0.17	0.17	0.15	0.57
1996	0.96	0.50	0.50	0.40	0.31	0.28	0.28	0.25	0.24	0.72
1997	1.35	0.55	0.55	0.45	0.39	0.35	0.30	0.31	0.29	1.05
1998	1.83	1.11	1.11	0.95	0.73	0.75	0.72	0.69	0.66	1.17
1999	2.75	1.03	1.03	0.81	0.50	0.45	0.41	0.36	0.34	2.41
2000	2.14	0.67	0.67	0.50	0.37	0.31	0.21	0.18	0.13	2.01
2001	1.57	0.70	0.70	0.31	-0.01	0.00	-0.12	-0.12	-0.11	1.68
2002	1.66	0.81	0.81	0.57	0.39	0.41	0.42	0.33	0.33	1.33
2003	0.83	0.23	0.23	0.20	0.14	0.15	0.13	0.11	0.11	0.72
2004	0.52	0.24	0.24	0.19	0.15	0.14	0.13	0.13	0.11	0.40
2005	0.45	0.21	0.21	0.17	0.15	0.13	0.14	0.12	0.12	0.33
2006	0.48	0.24	0.24	0.23	0.17	0.15	0.14	0.14	0.13	0.34
2007	0.52	0.18	0.18	0.10	0.08	0.07	0.05	0.05	0.04	0.47
2008	2.07	1.16	1.16	0.98	0.82	0.76	0.71	0.60	0.66	1.41
2009	0.52	0.05	0.05	-0.11	-0.07	-0.15	-0.18	-0.23	-0.20	0.72
2010	0.53	0.28	0.28	0.21	0.20	0.20	0.17	0.18	0.18	0.35
Mean	1.34	0.87	0.75	0.62	0.55	0.53	0.50	0.48	0.47	0.87

of holdings, the TNA and the selection spread, I take the low explanatory power in the all fund sample as further evidence that some funds load on unobservable systematic risks when stock picking, while others do not expose themselves to these risks.

How can the large cross sectional variation of the selection spread be explained if not by different idiosyncratic risk exposure? Table 6 shows the results of a cross sectional year by year regression of the selection spread on the annual risk loading of the fund's holdings portfolio on the four Carhart factors. This, at the same time, can serve as a test of the calibration of the MPPM. If these factors are indeed risks that lead to a reduction in certainty equivalent based on the utility model used in the MPPM, the cross sectional regression coefficients should be equivalent to the time series averages of the risk prices of these factors. The mean coefficient on the market factor suggests that the MPPM with $\rho = 2.7$ slightly, but insignificantly, underprices market risk (if the four factor model is correct) or that there are omitted risk factors in the model. The average size risk coefficient is only 11*bps* (insignificant) away from the average market return to the size strategy during the 28 year sample period. The momentum and value prices significantly diverge and even have different signs. Either, an investor with power utility of the type specified in section 3.1 is not averse to value and momentum risks (maybe because they are no risks) or there are rare value and momentum crashes which cannot be captured in this 28 year period. In light of the controversial risk story of the value and momentum factors, I consider the risk prices estimated as supporting the well-calibration of my model. On average 60% of the cross sectional variation of the selection spread can be explained by different exposures to the four Carhart factors and the failure of the DGTW model to precisely correct for that. The traditional Jensen one factor alpha by itself explains a lot less of the variation, with on average 76% of the variation explained by factors outside of the model. In the highest selection spread year (19987), which also had the highest cross sectional variation in terms of the difference of the top and the bottom decile portfolio, the Jensen model alone does not explain any of the variation while the 4 factor model captures 78% of it. If the value, size and momentum factors are no risk factors by themselves, they at least seem to be a decent proxy for the true risk factors. The lower mean risk price of only 4.13% in the Jensen model is a further indication that risk factors were

Table 6: Systematic risk as source of cross sectional variation of the selection spread.
Regression coefficients of the returns of the holdings portfolio on the 4 Carhart factors (left columns) or the the market factor only (right columns) are the regressors, the selection spread is the regressant. On average, the selection spread increases by 6.81% when the beta exposure increases by 1. The last line gives the mean annual market risk prices of the 4 factors over the 28 year period (from Kenneth French's web database). The selection spread and the risk prices are in %. The last line gives the difference between the mean cross sectional risk prices and the market price of risk, labeled by its significance. * signals significance at the 10% level, ** and *** at the 5% and the 1% level respectively.

	4 Factor Coefficients						1 Factor		
	Constant	Market	Value	Size	Momentum	R^2	Constant	Market	R^2
1983	-2.51***	3.29***	-2.16***	0.39*	0.45*	0.73	-1.46***	2.84***	0.38
1984	-8.08***	7.89***	-2.19***	0.45***	-3.75***	0.68	-5.47***	5.91***	0.33
1985	-4.71***	4.75***	-2.76***	0.97***	-2.76***	0.74	-2.03***	3.02***	0.22
1986	-5.09***	5.24***	-1.94***	0.43**	-0.40**	0.58	-0.72	1.74***	0.07
1987	-26.15***	27.03***	-6.29***	2.94***	8.35***	0.78	4.92*	0.09	0.00
1988	-1.82***	2.17***	-0.95***	0.95***	-1.19***	0.58	-0.26	1.25***	0.13
1989	-2.14***	2.54***	-0.41***	0.44**	0.46**	0.33	-2.03***	2.68***	0.24
1990	-9.74***	9.75***	-2.09***	1.52***	-0.91***	0.85	-9.04***	9.60***	0.70
1991	-6.81***	7.01***	-0.96***	1.73***	-0.16	0.78	-5.38***	6.27***	0.41
1992	-1.13***	1.45***	-0.46***	1.18***	0.94***	0.53	-1.47***	2.12***	0.26
1993	-0.87***	1.25***	-0.96***	0.55***	0.90***	0.67	0.48***	0.26**	0.01
1994	-2.53***	2.83***	-0.74***	0.37***	1.21***	0.59	-1.97***	2.49***	0.27
1995	-0.67***	0.96***	-0.67***	0.98***	0.34***	0.54	-0.65***	1.25***	0.21
1996	-3.18***	3.62***	-2.15***	1.47***	-0.29**	0.67	-2.10***	2.90***	0.18
1997	-3.33***	3.78***	-2.69***	1.24***	-0.66***	0.66	-0.12	1.18***	0.04
1998	-15.34***	16.11***	-6.57***	3.96***	-2.71***	0.75	-3.08***	5.13***	0.09
1999	-7.86***	10.22***	-5.11***	3.79***	0.81**	0.51	-2.26***	4.46***	0.09
2000	-3.90***	6.82***	-6.14***	2.97***	3.41***	0.36	-1.54***	3.64***	0.11
2001	-10.50***	12.25***	-6.73***	4.43***	-17.39***	0.73	-11.35***	11.68***	0.39
2002	-8.20***	9.42***	-3.00***	2.14***	-11.41***	0.78	-9.23***	10.33***	0.34
2003	-2.94***	3.23***	-0.61***	1.32***	-5.11***	0.43	-1.55***	2.04***	0.10
2004	-1.94***	2.14***	-0.27***	1.09***	1.75***	0.52	-1.19***	1.81***	0.20
2005	-1.88***	2.11***	-0.02	0.68***	0.64***	0.41	-1.14***	1.53***	0.23
2006	-0.89***	1.24***	-0.06	0.92***	0.45***	0.39	-0.42***	0.97***	0.13
2007	-3.66***	4.22***	0.21***	0.36***	0.22**	0.42	-3.82***	4.42***	0.38
2008	-16.08***	17.55***	2.60***	1.77***	-7.44***	0.55	-9.80***	11.53***	0.40
2009	-8.02***	9.02***	5.26***	0.62***	-16.50***	0.68	-0.59	1.31**	0.00
2010	-7.78***	8.07***	1.55***	0.67***	1.76***	0.67	-7.18***	7.65***	0.47
Mean	-5.99	6.64	-1.65	1.44	-1.75	0.60	-2.87	3.93	0.23
Risk Prices		7.47	4.45	1.32	7.11			7.47	
Dif		0.82	6.11**	-0.12	8.86***			3.53	

omitted.

Different loadings in the known risk factors explain large parts of the cross sectional variation of the selection spread. It could be, however, that all funds load very differently on the known risk factors, explaining the cross section differences in the spread, however, on average their exposure to these factors exactly resembles that of the benchmark portfolios. Than, the mean undiversifiable level of the selection spread would not be explained by the known risk factors. Table 7 gives insights into this question by looking at the difference of the risk exposures of the holdings and the benchmark porfolios. Funds, on average, have a beta 0.03 higher than their benchmark in the four factor model and 0.06 higher than their benchmark in the one factor model. Therefore, funds are systematically riskier than their benchmark in terms of the market factor. At the same time, funds have a 0.1 lower exposure to the value factor and a 0.02 higher exposure to each, the size and the momentum factors. In a four factor model world, funds expose themselves to less known systematic risk than their benchmark. In terms of known risk factors, this risk reduction increases their return to stock picking by 25*bps*. In the one factor model world, funds expose themselves to more systematic risks, which explains on average 51*bps* of their presumed stock picking skill.

Table 8 summarizes all components of the selection spread. On average, by stock picking, funds lose 132*bps* in certainty equivalent compared to their benchmark. Of these, 87*bps* are due to additional idiosyncratic risk which can be diversified away by holding a larger fund portfolio. Interestingly, funds are less risky than their benchmarks in terms of their Carhart four factor exposure. Their safer betting on these factors explains negative 25*bps* of the selection spread, leaving the remaining 73*bps* to unidentified but undiversifiable risk factors. In the Jensen one factor model, fund's higher exposure to the market factor almost fully explains the undiversifiable part of the selection spread, leaving only negative 5*bps* to unidentified systematic risk. From table 6 it is known however, that the one factor model does not perform very well in explaining the cross sectional variation of the selection spread. Therefore, omitted risk factors likely balance out in this sample and the one factor model. The large time series variation of the unidentified risk exposure even in the one factor model is further evidence that the different one factor exposure

Table 7: Difference in four factor Carhart and one factor Jensen risk loadings between the holdings portfolio and the actual benchmark portfolio.

Regression coefficients of the returns of the holdings portfolio on the 4 Carhart factors are the regressors, minus the regression coefficients of the same regression of the characteristic benchmark portfolio return. Last column gives the differences in systematic returns of both. If the benchmark completely adjusts for the four factors, the difference should be zero. All numbers are in %.

	4-Factor Regression Coefficients						**1-Factor**		
	Constant	Market	Value	Size	Momentum	Explained	Constant	Market	Explained
1983	0.01	0.02	-0.40	-0.40	0.13	-10.82	-0.00	0.08	0.95
1984	0.00	0.10	-0.15	-0.10	-0.00	-2.67	-0.00	0.17	-1.09
1985	-0.00	0.03	-0.15	0.09	0.01	0.89	-0.00	0.11	2.29
1986	-0.00	-0.03	-0.42	-0.07	0.17	-1.87	-0.00	0.05	0.29
1987	0.00	-0.04	-0.42	-0.07	0.17	2.58	0.00	0.07	0.08
1988	0.00	0.06	-0.18	0.07	0.07	-1.64	-0.00	0.10	0.97
1989	-0.00	0.23	-0.23	0.32	-0.11	-1.14	0.00	-0.03	-0.40
1990	0.00	0.07	-0.07	0.05	-0.01	-0.83	0.00	0.13	-1.36
1991	0.00	0.10	-0.13	0.02	-0.03	3.93	0.00	0.11	2.89
1992	0.00	0.06	-0.06	0.09	0.00	-0.30	-0.00	0.11	0.57
1993	0.00	-0.08	-0.20	-0.04	0.09	-2.05	0.00	0.09	0.82
1994	0.00	0.06	-0.13	0.08	-0.01	-0.03	0.00	0.05	-0.01
1995	-0.00	0.05	-0.12	0.07	-0.00	0.79	-0.00	0.11	2.97
1996	0.00	0.04	-0.02	0.05	0.01	0.66	0.00	0.04	0.76
1997	0.00	-0.00	-0.05	0.03	-0.01	-0.73	-0.00	0.02	0.32
1998	0.00	-0.00	-0.04	0.12	0.03	-1.22	-0.02	0.33	3.65
1999	0.00	0.06	0.13	0.07	0.07	0.66	0.00	0.03	0.61
2000	0.00	0.02	0.08	0.03	0.04	3.80	0.01	-0.01	0.77
2001	-0.00	0.01	0.06	0.04	-0.02	1.59	-0.00	-0.01	0.10
2002	-0.00	0.00	0.05	0.06	-0.06	-0.95	-0.00	0.05	-1.15
2003	-0.00	-0.02	0.05	0.14	0.02	2.22	0.00	-0.01	-0.16
2004	0.00	-0.04	-0.10	-0.08	0.14	-1.49	0.00	0.04	0.53
2005	0.00	0.07	0.07	-0.02	-0.05	0.27	0.00	0.04	0.28
2006	0.00	0.01	-0.06	0.02	0.03	-0.77	-0.00	0.07	0.71
2007	0.00	0.01	-0.00	0.00	0.00	0.19	0.00	0.01	0.18
2008	0.00	-0.02	-0.23	0.05	-0.02	0.50	0.00	0.04	-1.50
2009	0.00	-0.02	-0.04	0.05	-0.02	0.90	0.00	-0.01	-0.06
2010	-0.00	-0.00	-0.03	0.03	0.05	0.64	0.00	0.01	0.19
Mean	0.00	0.03	-0.10	0.02	0.02	-0.25	0.00	0.06	0.51

does not fully explain the selection spread. Whatever the exact nature of the undiversifiable risks is, funds — compared to their benchmarks — have excess exposure to them. This costs them, on average, 47*bps* in annual performance in stock picking. Further, funds are not well diversified, leaving the task of diversification to the investor. Total returns to stock picking before fees, but after risk adjustment, are 42*bps* in a diversified fund portfolio, and negative 44*bps* in a single fund portfolio in my slightly survivorship biased sample.

Table 8: Components of the selection spread.

The selection spread broken down into its risk components. Column 3 is the selection spread due to idiosyncratic risk. Columns 4 and 6 are the selection spread to to imprecise adjustments to the commonly used systematic risk factors in a Carhart or Jensen model, respectively. Columns 5 and 7 are quantify the undiversifiable risk components that are not explained by the commonly used systematic risk factors but by some higher order systematic risks. Columns 8 and 9 give the actual risk adjusted stock picking gains to the average investor in a 1 fund (8) and in a 30 fund (9) portfolio. Note that figures in the last two columns are subject to some survivorship bias. All numbers are in %.

		Idiosyncratic Risk	**Systematic Risk**				**MP Selection**	
			Carhart Model		Jensen Model		Adjustment	
	Selection Spread	Diversifiable	Identified	Unidentified	Identified	Unidentified	Undiversified	Diversified
1983	1.14	0.71	-10.82	11.25	0.95	-0.52	-1.55	-0.84
1984	1.43	0.78	-2.67	3.32	-1.09	1.74	-2.92	-2.14
1985	1.18	0.54	0.89	-0.25	2.29	-1.65	-0.74	-0.20
1986	1.12	0.58	-1.87	2.41	0.29	0.25	-1.83	-1.26
1987	5.01	1.32	2.58	1.10	0.08	3.61	-3.04	-1.71
1988	1.05	0.73	-1.64	1.96	0.97	-0.65	-1.75	-1.01
1989	0.97	0.56	-1.14	1.55	-0.40	0.81	0.40	0.96
1990	2.28	0.88	-0.83	2.23	-1.36	2.77	0.34	1.22
1991	1.84	0.79	3.93	-2.88	2.89	-1.85	0.32	1.11
1992	0.84	0.55	-0.30	0.59	0.57	-0.28	-0.40	0.15
1993	0.72	0.58	-2.05	2.19	0.82	-0.68	0.23	0.81
1994	0.63	0.55	-0.03	0.11	-0.01	0.10	1.37	1.91
1995	0.71	0.57	0.79	-0.64	2.97	-2.83	-0.82	-0.25
1996	0.93	0.72	0.66	-0.44	0.76	-0.54	0.18	0.90
1997	1.02	1.05	-0.73	0.70	0.32	-0.35	-2.51	-1.46
1998	1.92	1.17	-1.22	1.97	3.65	-2.90	-1.53	-0.36
1999	2.45	2.41	0.66	-0.61	0.61	-0.57	-0.42	1.99
2000	2.60	2.01	3.80	-3.21	0.77	-0.17	4.62	6.62
2001	1.67	1.68	1.59	-1.60	0.10	-0.12	-1.50	0.18
2002	1.63	1.33	-0.95	1.24	-1.15	1.45	-2.63	-1.30
2003	0.63	0.72	2.22	-2.31	-0.16	0.07	1.00	1.72
2004	0.54	0.40	-1.49	1.63	0.53	-0.39	0.46	0.86
2005	0.44	0.33	0.27	-0.16	0.28	-0.17	1.02	1.35
2006	0.53	0.34	-0.77	0.96	0.71	-0.52	-1.33	-0.99
2007	0.48	0.47	0.19	-0.19	0.18	-0.18	1.59	2.06
2008	1.96	1.41	0.50	0.05	-1.50	2.05	-2.04	-0.63
2009	0.61	0.72	0.90	-1.00	-0.06	-0.05	1.40	2.12
2010	0.47	0.35	0.64	-0.52	0.19	-0.07	-0.36	-0.01
Mean	1.31	0.87	-0.25	0.69	0.51	-0.06	-0.44	0.42

5.3 Analysis of selection risks in the time series

Time series variation of all selection risks (the equally weighted selection spread from table 2), the diversifiable selection risk (the difference between the 1 fund and the 30 fund portfolio selection spread from table 5) and the undiversifiable (30 fund portfolio) selection spread is analyzed in this section. The hypothesis is that the time series variation is largely due to different capturability of higher order systematic risks in each year. Some variation of systematic risk might also come from herding, as diversification benefit will be lower in high herding years. I therefore expect a positive coefficient of the herding variable in the systematic risk regression, and a negative coefficient in the idiosyncratic risk regression. Time series variation of herding, however, is low.

It is well known that especially kurtosis can only be reliably estimated in extremely large samples. For instance, Bai and Ng (2005) show that, for some distributions, kurtosis is underestimated even with a sample size of 5000 observations. The annual MPPM measure relies on twelve observations. The brevity of the sample could lead to underestimation of the costs of loading on higher order systematic risks in most years, and overestimation of these costs in other years. Consider, for example, a fund manager who loads on some type of higher order risk that is characterized by extreme, but very rare, negative payoffs (what the banking jargon refers to as "black swan events") and regular, but low, positive payoffs[26]. While this strategy pays — in the long term — just the market price of risk, in the short term it might be mistaken for outperformance most of the time. At the same time, in periods where the "black swan event" occurs, this strategy will accidentally be marked "underperform". The same could be true for higher order systematic risks taken by some fund managers. Without attempting to draw conclusions on the exact type of these unidentified systematic risks, I proxy for their capturability with the second (annualized) to forth moment of the daily returns of the CRSP value weighted index every year. Especially high kurtosis years can be considered a measure of the occurrence of such "black swan events". The argument is that in high skewness or kurtosis years, the true fund exposure to tail risks is uncovered. The certainty equivalent measure might overestimate the damage done by higher moment exposure in high moment years like, for example, 1987. At

[26] E.g. a deep out-of-the-money short option strategy.

the same time, in a low kurtosis year like 2004, we cannot observe the moment exposure of the fund, because there are no extreme realizations. Looking at the 1987 data, it seems as if funds load heavily on tail risks.

Table 9 analyses the time series variation of the selection spread. It can be seen that the herding measure has some, but very low, explanatory power of the systematic component of the selection spread if the standard deviation of the CRSP value weighted index is controlled for. It does not explain any of the diversifiable risk. The moment realizations of the CRSP value weighted index can explain 74% of the time series variation of the average systematic component of the selection spread, with kurtosis significant at the 1% level and skewness significant at the 10% level. Parts of the level of diversifiable risks can be explained by the standard deviation of the CRSP value weighted index. Low diversification is especially costly in high volatility years.

Table 9: The selection spread in a time series regression.

	Total	Total	Total	Diversifiable	Diversifiable	Diversifiable	Systematic	Systematic	Systematic
maxherding	4.98	6.12	2.24	-0.45	0.20	0.33	5.44	5.88*	1.81
	(1.22)	(1.77)	(0.77)	(-0.20)	(0.11)	(0.16)	(1.87)	(2.07)	(1.02)
CRSP VW Standard Dev.		6.88**	3.56		3.93**	3.84**		2.67	-0.68
		(3.34)	(1.87)		(3.54)	(2.87)		(1.58)	(-0.59)
CRSP VW Skewness			0.58			0.09			0.53*
			(1.54)			(0.34)			(2.33)
CRSP VW Kurtosis			0.13**			0.01			0.13***
			(2.92)			(0.21)			(4.76)
Constant	-0.33	-1.80	-0.66	1.02	0.18	0.14	-1.33	-1.90	-0.70
	(-0.24)	(-1.46)	(-0.65)	(1.36)	(0.27)	(0.19)	(-1.37)	(-1.88)	(-1.14)
R^2	0.05	0.35	0.62	0.00	0.33	0.34	0.12	0.20	0.74

5.4 Risks incured by market timing

Figure 2: Equal weighted DGTW and MP timing measures 1983–2010.

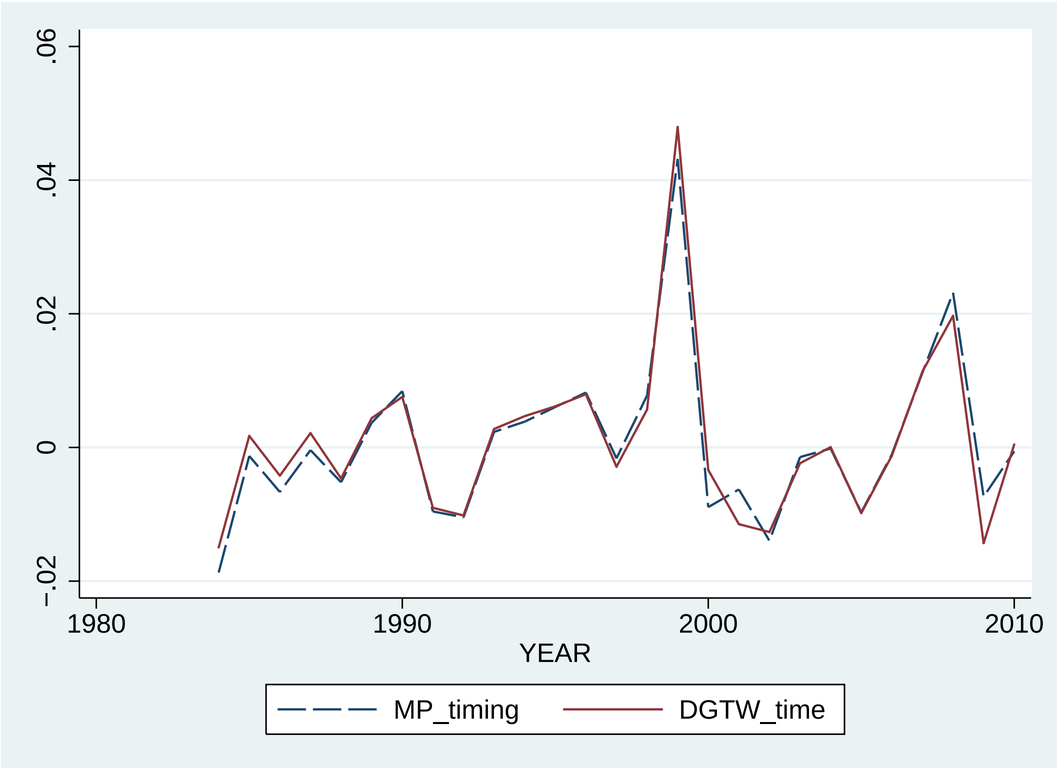


Figure 2 maps the timing measure — in certainty equivalent terms — against the classic DGTW timing measure. As expected, both are almost identical over the entire sample period. The difference is statistically insignificant $3bps$ on an equally weighted basis and statistically insignificant $-2bps$ value weighted. Fund managers do not systematically move to higher- or lower-risk characteristic portfolios over time. Any spread between the two measures would be largely attributable to differences in systematic risk, as all characteristic benchmarks are equally well diversified. Managers could only be manipulating this measure if systematic risk characteristics were changing over time and managers were able to anticipate these changes. If systematic risk characteristics are constant, managers would have traded into the riskiest portfolio after, the latest, 125 periods and further excessive risk taking would not be possible.

In table 10, I analyse the cross sectional and time series variation of the timing spread. Also there is some cross sectional variation, this is not persistent (the average difference between

Table 10: Difference between the top and the bottom decile timing spread each year and persistence of the difference (in %).

The first two columns give the average, equal weighted, DGTW- and MP timing measures in each year. Column 3 gives the difference between the decile of funds with the highest timing spread and the decile of funds with the lowest timing spread. Columns 4 to 9 check the persistence of this difference. Therefore, e.g. in column 8 (t=4) in year 1983, the difference between the decile of funds that had the highest and the lowest timing spread in 1983 is given for 1987. The last column reports the value weighted results in $t = 0$

	MP Time	DGTW Time	Time Spread	Decile 10 - Decile 1						VW
	t=0	t=0	t=0	t=0	t=1	t=2	t=3	t=4	t=5	t=0
1984	-1.87	-1.50	0.37	1.68	1.08	0.18	0.56	0.01	0.15	0.34
1985	-0.13	0.17	0.30	1.54	0.60	-0.55	0.02	0.01	-0.16	0.22
1986	-0.67	-0.42	0.24	1.63	-0.20	-0.20	0.10	0.23	0.10	0.27
1987	-0.04	0.21	0.25	5.43	0.19	-0.03	0.19	-0.08	0.07	0.29
1988	-0.52	-0.46	0.06	1.10	0.05	-0.25	0.74	-0.01	0.14	0.07
1989	0.37	0.44	0.07	1.09	0.62	0.95	0.03	0.13	0.29	0.12
1990	0.84	0.76	-0.09	2.40	0.36	-0.08	0.21	0.00	-0.10	-0.07
1991	-0.96	-0.90	0.05	1.75	-0.13	0.05	0.40	0.03	0.01	0.06
1992	-1.05	-1.02	0.03	1.33	0.20	0.26	-0.32	-0.64	-0.11	-0.03
1993	0.23	0.28	0.04	0.95	0.14	0.00	-0.00	-0.61	-0.66	0.06
1994	0.39	0.47	0.08	0.83	-0.12	-0.48	-0.29	0.69	2.88	0.09
1995	0.60	0.62	0.01	0.77	-0.07	-0.14	-0.01	0.33	-0.37	0.03
1996	0.82	0.80	-0.03	1.30	-0.16	0.00	-0.02	-1.13	-3.43	-0.04
1997	-0.16	-0.29	-0.13	1.90	-0.22	0.55	-1.24	-0.89	-0.04	-0.15
1998	0.78	0.57	-0.21	3.14	-0.72	-2.63	-0.48	-0.05	0.07	-0.19
1999	4.32	4.80	0.48	4.99	3.16	-0.86	-0.26	-0.10	0.07	0.32
2000	-0.89	-0.34	0.56	9.70	2.94	0.92	0.29	-0.02	0.07	0.12
2001	-0.63	-1.15	-0.52	9.35	0.75	0.31	-0.03	0.00	0.02	-1.02
2002	-1.40	-1.27	0.13	3.80	0.56	0.02	0.01	-0.06	-0.01	0.12
2003	-0.14	-0.24	-0.09	1.22	0.14	0.06	0.04	0.08	0.10	-0.17
2004	-0.02	0.00	0.02	0.86	0.11	-0.13	0.21	-0.29	0.04	-0.02
2005	-0.97	-0.98	-0.01	0.71	0.03	-0.05	0.56	0.42	0.31	-0.03
2006	-0.10	-0.12	-0.02	0.75	-0.05	0.75	0.13	0.07		-0.03
2007	1.12	1.14	0.02	1.02	-0.51	-0.01	0.02			0.04
2008	2.32	1.97	-0.35	4.77	2.27	0.21				-0.22
2009	-0.74	-1.43	-0.70	4.65	0.34					-0.71
2010	-0.06	0.05	0.10	2.12						0.06
Mean	0.05	0.08	0.03	2.62	0.44	-0.05	0.04	-0.08	-0.03	-0.02

decile 10 and decile 1 is already insignificant in t=1) and therefore not a characteristic of the fund. Of course, at times managers might move to slightly riskier or less risky benchmarks when reallocating assets. However, there do not seem to be managers who consistently move to riskier benchmarks while others consistently move to less risky benchmarks. Also, there is some, but little, time series variation. I conclude that the DGTW timing measure is sufficient from a risk correction perspective. Therefore market timing is not discussed further in this paper.

6 Conclusion

Stock picking increases the risk of the portfolio, and the returns to stock picking are reduced if this risk increase is adjusted for. The total risk induced by stock picking lowers certainty equivalent returns by 131*bps* in an equally weighted sample and by 94*bps* in a value weighted sample. A large part of this additional risk is explained by poor diversification of the stock picking portfolios. Investors can avoid this additional idiosyncratic risk by holding larger, multi-fund portfolios. I can show that, even after diversifying, the additional risk in stock picking portfolios reduces the certainty equivalent return to stock picking by almost 0.5% per year. This cost comes in addition to other costs to active management discussed in the literature, such as fees and transaction costs, and pushes the total, after fee returns to stock picking further below zero. Funds do not uniformly expose themselves to this type of risk. While some funds persistently lose certainty equivalent by stock picking, others are persistently less risky than their benchmark portfolio. The high cross sectional variation of the risk exposure in the stock picking portfolio can, in part, be explained by different loadings on the market factor and on the value-, size-, and momentum factors. The remainder is most likely explained by different exposure to other, unidentified risk types. Funds, on average, exposed themselves slightly less to the market-, value-, size- and momentum factors than their benchmark portfolio, increasing their risk adjusted returns to stock picking by 25*bps* if the four factor model is correct. Unidentified, undiversifiable risks, however, reduce their risk adjusted returns to stock picking by 69*bps*. The time series variation of the selection risk can be largely explained by different capturability of the risk in different years. Market timing portfolios are not riskier than their benchmarks. Further research should make an effort to learn more about the undiversifiable, but unidentified risks that cost the investor up to 69*bps* in certainty equivalent returns per annum.

7 Appendix

Table 11: The average value weighted selection spread of multi fund portfolios in each year (in %).

Multi fund portfolios are build as described in section 3.4. MP selection measures are based on the portfolio certainty equivalent. The last column gives the diversification gain as the difference between the selection spread of the 30 fund portfolio and the 1 fund portfolio.

	No. of Funds in Portfolio									**Divers. Gain**
	1	2	3	5	8	10	15	20	30	30-1
1983	0.86	0.67	0.67	0.56	0.52	0.48	0.46	0.43	0.44	0.42
1984	1.12	0.92	0.92	0.81	0.80	0.77	0.76	0.73	0.73	0.39
1985	1.07	0.86	0.86	0.79	0.76	0.77	0.71	0.72	0.72	0.35
1986	0.81	0.66	0.66	0.61	0.57	0.55	0.51	0.50	0.48	0.33
1987	4.84	4.62	4.62	4.22	4.34	4.12	4.09	4.08	4.05	0.80
1988	0.80	0.65	0.65	0.53	0.50	0.49	0.46	0.39	0.39	0.41
1989	0.75	0.60	0.60	0.52	0.48	0.48	0.46	0.42	0.40	0.35
1990	1.39	1.42	1.42	1.24	1.15	1.14	1.17	1.09	1.12	0.27
1991	1.40	1.25	1.25	1.14	1.06	1.13	0.99	1.04	1.05	0.35
1992	0.63	0.64	0.64	0.45	0.38	0.39	0.30	0.29	0.27	0.37
1993	0.64	0.37	0.37	0.33	0.21	0.23	0.19	0.16	0.16	0.47
1994	0.51	0.47	0.47	0.41	0.39	0.36	0.32	0.33	0.31	0.20
1995	0.72	0.50	0.50	0.49	0.32	0.33	0.36	0.34	0.30	0.42
1996	0.75	0.59	0.59	0.57	0.65	0.46	0.44	0.39	0.36	0.39
1997	1.38	0.86	0.86	0.83	0.74	0.57	0.47	0.49	0.45	0.93
1998	0.89	1.10	1.10	0.82	0.69	0.83	0.72	0.80	0.62	0.27
1999	1.28	1.15	1.15	1.17	0.81	0.66	0.72	0.70	0.56	0.72
2000	2.43	1.23	1.23	0.73	0.86	0.66	0.46	0.33	0.28	2.15
2001	0.96	1.10	1.10	0.88	0.40	0.21	0.21	0.10	0.09	0.86
2002	0.71	0.75	0.75	0.47	0.33	0.39	0.31	0.32	0.20	0.51
2003	0.15	0.10	0.10	0.12	0.07	0.03	0.05	0.03	0.02	0.14
2004	0.27	0.27	0.27	0.16	0.14	0.12	0.10	0.11	0.08	0.19
2005	0.17	0.13	0.13	0.10	0.09	0.10	0.10	0.07	0.07	0.10
2006	0.22	0.23	0.23	0.19	0.14	0.14	0.11	0.12	0.10	0.12
2007	0.33	0.15	0.15	0.11	0.08	0.08	0.05	0.05	0.05	0.28
2008	1.00	1.12	1.12	0.59	0.62	0.60	0.58	0.52	0.42	0.58
2009	0.12	0.14	0.14	-0.03	-0.10	-0.14	-0.21	-0.15	-0.22	0.34
2010	0.33	0.31	0.31	0.25	0.23	0.27	0.18	0.22	0.20	0.14
Mean	0.95	0.84	0.82	0.68	0.61	0.58	0.54	0.52	0.49	0.46

Table 12: The average selection spread sorted by the lagged Selection Spread decile each year (in %)

	1A	1B	1C	1	2	3	4	5	6	7	8	9	10	10A	10B	10C	All	10-1	10C-1A
1984	-0.04	0.65	1.22	0.53	1.35	0.41	1.13	0.50	1.04	1.36	1.59	2.29	2.87	1.84	3.16	3.60	1.43	2.33	3.65
1985	-0.06	0.20	1.32	0.49	0.34	0.49	0.37	0.71	0.76	0.93	0.97	1.28	4.31	1.38	3.20	9.31	1.18	3.82	9.36
1986	-0.08	0.21	1.81	0.64	0.21	0.43	0.79	1.27	0.85	0.62	1.19	1.51	3.22	0.63	2.87	6.15	1.12	2.57	6.23
1987	0.63	4.69	8.43	4.59	3.68	3.40	3.83	3.36	5.44	8.04	4.43	9.65	10.28	1.79	9.46	21.90	5.01	5.69	21.27
1988	-0.11	0.71	1.30	0.59	0.52	0.54	0.72	0.54	0.75	1.17	0.84	2.04	2.41	0.79	2.08	4.85	1.05	1.83	4.96
1989	0.15	0.48	0.93	0.50	0.42	0.59	0.73	0.72	0.63	1.33	1.14	1.50	1.84	0.40	1.99	3.35	0.97	1.34	3.20
1990	-1.43	0.52	4.85	1.31	1.10	0.96	1.44	1.28	2.68	2.15	2.46	4.54	6.31	3.27	6.43	9.94	2.28	4.99	11.37
1991	-1.17	-0.13	1.37	-0.06	0.48	0.49	1.03	1.21	1.16	1.46	2.63	2.17	4.70	2.83	4.82	6.84	1.84	4.76	8.01
1992	-0.35	0.08	0.71	0.15	0.13	0.20	0.66	0.66	0.68	1.02	1.21	1.90	3.03	1.77	2.88	4.66	0.84	2.88	5.00
1993	-0.10	0.11	0.42	0.14	0.25	0.24	0.28	0.34	0.45	0.53	1.01	1.43	2.11	1.06	2.21	3.23	0.72	1.97	3.33
1994	-0.24	0.06	0.38	0.06	0.08	0.17	0.41	0.27	0.23	0.92	0.86	1.40	3.14	1.39	2.44	5.59	0.63	3.08	5.83
1995	-0.18	0.15	0.61	0.18	0.09	0.19	0.23	0.39	0.37	0.68	0.74	1.25	2.11	1.01	1.88	3.55	0.71	1.93	3.73
1996	-0.50	0.14	0.77	0.12	0.31	0.18	0.25	0.34	0.39	0.72	1.07	1.94	3.33	0.99	2.39	6.77	0.93	3.21	7.26
1997	-0.79	0.03	0.84	0.03	0.07	0.15	0.48	0.46	0.77	0.72	1.16	1.95	4.02	1.03	3.45	7.84	1.02	3.99	8.63
1998	-4.57	-1.34	2.38	-1.23	0.11	0.37	-0.04	0.42	1.62	1.97	3.30	4.66	7.96	3.20	7.80	13.12	1.92	9.19	17.70
1999	-0.63	0.36	2.94	0.86	0.46	0.96	0.82	1.36	1.24	2.02	2.86	4.80	7.94	1.65	5.71	16.76	2.45	7.07	17.39
2000	-4.28	0.01	3.46	-0.32	0.42	0.65	1.56	1.12	1.41	1.14	3.74	5.45	11.06	1.04	9.68	23.03	2.60	11.38	27.31
2001	-3.42	-1.17	2.39	-0.77	0.10	-0.30	-0.28	0.23	0.15	0.73	0.59	2.79	10.71	-0.37	6.62	26.32	1.67	11.49	29.75
2002	-1.66	-0.17	1.98	0.03	0.11	0.20	0.20	0.38	0.29	0.82	0.97	2.41	9.90	1.65	7.88	20.17	1.63	9.87	21.83
2003	-0.75	-0.21	0.73	-0.08	0.02	0.05	0.10	0.20	0.21	0.40	0.64	0.77	3.07	0.88	2.01	6.31	0.63	3.15	7.06
2004	-0.31	0.06	0.52	0.09	0.09	0.07	0.11	0.13	0.25	0.31	0.47	0.85	2.89	0.77	2.32	5.57	0.54	2.80	5.89
2005	-0.32	-0.01	0.62	0.09	0.10	0.09	0.11	0.17	0.21	0.26	0.40	0.82	2.43	0.54	1.63	5.19	0.44	2.34	5.51
2006	-0.34	0.04	0.57	0.09	0.14	0.14	0.16	0.21	0.28	0.43	0.62	0.88	2.09	0.60	1.60	4.13	0.53	2.00	4.47
2007	-0.20	0.07	0.69	0.19	0.15	0.17	0.20	0.31	0.27	0.39	0.54	0.74	1.68	0.20	1.12	3.77	0.48	1.49	3.97
2008	-1.66	0.07	3.73	0.68	0.07	0.35	0.62	1.06	1.22	1.66	2.04	2.31	9.42	1.26	7.47	19.52	1.96	8.73	21.18
2009	-2.55	-0.66	1.49	-0.60	0.18	0.24	-0.08	0.07	0.34	0.02	0.63	1.74	4.17	-0.54	2.75	10.52	0.61	4.77	13.07
2010	-1.70	-0.03	1.03	-0.23	0.05	0.38	0.33	0.26	0.21	0.41	0.61	0.75	1.88	-0.22	1.52	4.41	0.47	2.11	6.11
Mean	-0.99	0.18	1.76	0.30	0.41	0.44	0.60	0.67	0.89	1.19	1.43	2.36	4.77	1.14	3.98	9.50	1.32	4.47	10.48

Table 13: Idiosyncratic risk exposure in the cross section .

Sel spread is in % and is regressed on the number of funds in the portfolio and the total known assets of the portfolio in each year. Distinction is made between different type of investment styles. The table is continued on the next page.

	All			Value			Growth			Small			Large		
	No.	TNA	R2	No.	TNA	R2	No.	TNA	R2	No.	TNA	R2	No.	TNA	R2
1983	-0.00	-0.00	0.03	0.00	-0.00	1.00	-0.01	0.00	0.04						
1984	-0.00	-0.00	0.02	-0.07	0.00	0.61	-0.01	-0.00	0.02						
1985	-0.00	-0.00	0.02	0.11	0.00	0.96	0.00	0.00	0.01						
1986	-0.00	-0.00	0.03	0.02	0.00	0.11	0.00	-0.00	0.00	0.00	-0.00	1.00			
1987	-0.01	0.00	0.01	0.32	-0.00	0.47	-0.01	0.00	0.07	0.00	-0.00	1.00			
1988	-0.00*	-0.00	0.02	-0.00	0.00	0.25	-0.02***	0.00	0.18	0.02	-0.00	1.00			
1989	-0.00**	-0.00	0.02	-0.00	-0.00	0.10	-0.01	0.00	0.03	-0.02	0.00	0.58			
1990	-0.00	-0.00	0.02	-0.00	0.00	0.04	-0.01	-0.00	0.03	-0.01	0.00	0.18			
1991	-0.00**	-0.00	0.02	0.00	0.00	0.01	-0.01**	0.00**	0.09	-0.01	0.00	0.32			
1992	-0.00***	-0.00	0.03	-0.00	0.00	0.00	-0.00	0.00	0.01	-0.00	0.00	0.15			
1993	-0.00***	-0.00	0.04	-0.00	0.00	0.02	-0.00**	0.00	0.04	-0.00**	0.00**	0.43			
1994	-0.00***	0.00	0.02	-0.00	-0.00	0.01	-0.00	0.00	0.01	-0.00**	0.00**	0.20	0.02	0.00	1.00
1995	-0.00***	-0.00	0.02	-0.00*	-0.00	0.06	-0.00	-0.00	0.01	-0.00**	0.00	0.08	-0.00	0.00	0.06
1996	-0.00**	-0.00	0.01	-0.00	0.00	0.04	-0.00	-0.00	0.01	-0.00*	0.00	0.04	-0.00	-0.00	0.18
1997	-0.00***	-0.00	0.01	-0.00	0.00	0.01	-0.00	-0.00	0.01	-0.00	-0.00	0.02	-0.00	-0.00	0.05
1998	-0.00***	-0.00	0.01	-0.00	0.00	0.01	-0.00	-0.00*	0.01	-0.00**	0.00	0.04	-0.00	-0.00	0.02
1999	-0.00***	-0.00	0.02	-0.00	-0.00	0.01	-0.01*	-0.00*	0.02	-0.00***	-0.00	0.05	-0.01	-0.00	0.08
2000	-0.01***	-0.00	0.02	-0.01*	0.00	0.03	-0.01**	-0.00	0.02	-0.01**	0.00	0.02	-0.01	-0.00	0.05
2001	-0.00***	-0.00	0.01	-0.00	0.00	0.00	-0.01**	-0.00	0.02	-0.00**	0.00	0.03	-0.00	0.00	0.03
2002	-0.00***	-0.00	0.01	-0.00	-0.00	0.01	-0.00	-0.00	0.01	-0.00**	0.00	0.02	-0.00	-0.00	0.01
2003	-0.00***	-0.00*	0.02	-0.00**	-0.00	0.01	-0.00**	-0.00	0.02	-0.00**	-0.00	0.02	-0.00**	0.00	0.03
2004	-0.00***	-0.00	0.02	-0.00	0.00	0.01	-0.00**	-0.00	0.02	-0.00**	-0.00	0.02	-0.00***	0.00	0.06
2005	-0.00***	-0.00	0.01	-0.00	-0.00	0.00	-0.00***	-0.00	0.02	-0.00**	-0.00	0.01	-0.00**	0.00**	0.05
2006	-0.00***	-0.00	0.03	-0.00***	0.00	0.04	-0.00***	-0.00**	0.03	-0.00***	-0.00	0.03	-0.00***	0.00	0.07
2007	-0.00***	-0.00	0.02	-0.00**	0.00	0.02	-0.00***	-0.00	0.03	-0.00***	-0.00	0.04	-0.00**	0.00**	0.06
2008	-0.00***	-0.00	0.03	-0.01***	0.00	0.05	-0.00	-0.00	0.01	-0.00***	0.00	0.05	-0.00	-0.00	0.03
2009	-0.00***	-0.00	0.01	-0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	0.02
2010	-0.00***	0.00	0.01	-0.00	-0.00	0.00	-0.00	0.00	0.01	-0.00	0.00	0.01	-0.00**	0.00	0.04
Mean	-0.00	-0.00	0.02	0.01	-0.00	0.14	-0.00	0.00	0.03	-0.00	0.00	0.21	-0.00	0.00	0.11

ctd.

	Tech			Financials			Index			Others		
	No.	TNA	R2	No.	TNA	R2	No.	TNA	R2	No.	TNA	R2
1983	0.08	-0.00	1.00							-0.00	-0.00	0.03
1984										-0.01	-0.00	0.02
1985	0.19	-0.00	0.38	-1.44	-0.00	1.00				-0.01	-0.00	0.04
1986	0.30	-0.00	0.87	0.01	-0.00	0.20				-0.01*	-0.00	0.05
1987	-0.42	0.00	0.13	1.02	-0.00	0.49	0.00	0.00	1.00	0.00	-0.00	0.00
1988	-0.09	0.00	0.05	0.14	-0.00	0.34	-0.00	-0.00	0.23	-0.00	-0.00	0.02
1989	0.01	0.00	1.00	0.03	-0.00	0.16	0.00	0.00	0.14	-0.00	0.00	0.02
1990	1.59**	0.00	1.00	0.36	0.00	0.92	-0.00	-0.00	0.03	-0.00	-0.00	0.01
1991	0.35	0.00	0.12	-0.38	-0.00	1.00	-0.00	-0.00	0.08	-0.00	-0.00	0.01
1992	0.19	-0.00	0.52	0.10	0.00	1.00	-0.00	-0.00	0.05	-0.00***	-0.00	0.05
1993	0.09	-0.00	0.72				-0.00	-0.00	0.18	-0.00**	-0.00	0.05
1994	-0.05	0.00	0.19				-0.00	-0.00	0.09	-0.00***	0.00	0.05
1995	-0.10	-0.00	1.00	0.00	-0.00	1.00	-0.00	-0.00	0.09	-0.00**	0.00	0.02
1996	-0.02	0.00	1.00	0.01	0.00	1.00	-0.00***	-0.00	0.30	-0.00	0.00	0.01
1997	-0.06	-0.00	0.22	0.08	0.00	0.75	-0.00	-0.00	0.00	-0.00**	0.00	0.01
1998	-0.16	-0.00	0.20	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.01	0.00	0.01
1999	-0.32*	-0.00	0.24	0.02	0.00	0.09	-0.00	-0.00	0.02	-0.01**	-0.00	0.01
2000	-0.21**	-0.00	0.34	0.03	-0.00	0.01	-0.00	-0.00	0.01	-0.01***	-0.00	0.02
2001	-0.31***	-0.00	0.26	0.00	0.00	0.01	-0.00	-0.00	0.04	-0.01*	-0.00	0.01
2002	-0.08*	-0.00	0.07	0.00	0.00	0.03	-0.00**	-0.00	0.06	-0.00*	-0.00	0.01
2003	-0.03**	-0.00	0.10	-0.00	0.00	0.03	-0.00**	-0.00	0.07	-0.00***	-0.00	0.02
2004	-0.00	-0.00	0.01	-0.01**	0.00	0.25	-0.00*	-0.00	0.03	-0.00***	-0.00	0.03
2005	-0.00	-0.00	0.01	0.00	0.00	0.13	-0.00**	-0.00	0.04	-0.00***	-0.00	0.03
2006	-0.00	-0.00	0.07	-0.00	0.00	0.12	-0.00	-0.00	0.02	-0.00***	-0.00	0.05
2007	-0.00	0.00	0.00	-0.01	-0.00*	0.25	-0.00*	-0.00	0.04	-0.00***	-0.00	0.04
2008	-0.00	-0.00	0.01	0.01	-0.00	0.24	-0.00	-0.00	0.03	-0.01***	-0.00	0.06
2009	0.00	-0.00	0.01	0.02	-0.00	0.34	0.00	-0.00	0.00	-0.01***	-0.00	0.04
2010	0.00	0.00	0.11	0.05	-0.00	0.17	0.00	-0.00	0.01	-0.00***	-0.00	0.03
Mean	0.03	0.00	0.36	0.00	-0.00	0.40	-0.00	0.00	0.11	-0.00	-0.00	0.03

References

Agarwal, V., G. Bakshi, and J. Huij (2009). Do higher-moment equity risks explain hedge fund returns? *Robert H. Smith School Research Paper No. RHS 06-153*.

Alexander, G. J., G. Cici, and S. Gibson (2007). Does motivation matter when assessing trade performance? an analysis of mutual funds. *The Review of Financial Studies 20*(1), pp. 125–150.

Antypas, A. (2009). *Selectivity, market timing and the morningstar star-rating system.* CESifo, Center for Economic Studies & Ifo Institute for economic research.

Aragon, G. and W. Ferson (2007). *Portfolio performance evaluation*, Volume 2. Now Publishers Inc.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*(1), 49–60.

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

Carhart, M. M. (1992). Persistence in mutual fund performance re-examined. *Working paper Graduate School of Business, University of Chicago, Chicago, Ill. Graduate School of Business, University of Chicago, Chicago, Ill.*

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance 52*(1), pp. 57–82.

Chung, Y., H. Johnson, and M. Schill (2006). Asset pricing when returns are nonnormal: Fama-french factors versus higher-order systematic comoments*. *The Journal of Business 79*(2), 923–940.

Cremers, K. and A. Petajisto (2009). How active is your fund manager? a new measure that predicts performance. *Review of Financial Studies 22*(9), 3329–3365.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance 52*(3), pp. 1035–1058.

Daniel, K. and S. Titman (1997). Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance 52*(1), pp. 1–33.

Daniel, K., S. Titman, and K. Wei (2001). Explaining the cross-section of stock returns in japan: Factors or characteristics? *The Journal of Finance 56*(2), 743–766.

Del Guercio, D. and P. Tkac (2008). Star power: The effect of morningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis 43*(4), 907–936.

Fama, E. and K. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*(1), 3–56.

Fama, E. and K. French (2010). Luck versus skill in the cross-section of mutual fund returns. *The*

Journal of Finance 65(5), 1915–1947.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies 20*(5), 1503–1546.

Goetzmann, W. and A. Kumar (2008). Equity portfolio diversification. *Review of Finance 12*(3), 433–463.

Hendricks, D., J. Patel, and R. Zeckhauser (1993). Hot hands in mutual funds: Short-run persistence of relative performance, 1974-1988. *Journal of Finance*, 93–130.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of finance*, 65–91.

Jensen, M. (1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *The Journal of Business 42*(2), 167–247.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263–291.

Lakonishok, J., A. Shleifer, and R. Vishny (1992). The impact of institutional trading on stock prices. *Journal of financial economics 32*(1), 23–43.

Lehmann, B. (1990). Residual risk revisited. *Journal of Econometrics 45*(1), 71–97.

Liew, J. and M. Vassalou (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics 57*(2), 221–245.

Malkiel, B. and Y. Xu (2002). Idiosyncratic risk and security returns. *University of Texas at Dallas (November 2002)*.

Merton, R. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867–887.

Morey, M. and A. Gottesman (2006). Morningstar mutual fund ratings redux. *Journal of Investment Consulting 8/1*.

Morningstar (2007). The morningstar rating methodology.

O'Neal, E. (1997). How many mutual funds constitute a diversified mutual fund portfolio? *Financial Analysts Journal*, 37–46.

Scott, R. and P. Horvath (1980). On the direction of preference for moments of higher order than the variance. *The Journal of Finance 35*(4), 915–919.

Sharpe, W. (1966). Mutual fund performance. *The Journal of Business 39*(1), 119–138.

Sias, R. (2004). Institutional herding. *Review of Financial Studies 17*(1), 165–206.

Treynor, J. and F. Black (1973). How to use security analysis to improve portfolio selection. *The Journal* *of Business 46*(1), 66–86.

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance 55*(4), 1655–1703.