

Swade, Alexander; Köchling, Gerrit; Posch, Peter N.

**Article — Published Version**

## Managerial behavior in fund tournaments—the impact of TrueSkill

Journal of Asset Management

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Swade, Alexander; Köchling, Gerrit; Posch, Peter N. (2021) : Managerial behavior in fund tournaments—the impact of TrueSkill, Journal of Asset Management, ISSN 1479-179X, Palgrave Macmillan UK, London, Vol. 22, Iss. 1, pp. 62-75, <https://doi.org/10.1057/s41260-020-00198-7>

This Version is available at:

<https://hdl.handle.net/10419/287611>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Managerial behavior in fund tournaments—the impact of TrueSkill

Alexander Swade<sup>1,2</sup> · Gerrit Köchling<sup>3</sup> · Peter N. Posch<sup>3</sup>

Revised: 3 November 2020 / Accepted: 22 November 2020 / Published online: 9 January 2021  
© The Author(s) 2021

## Abstract

Measuring mutual fund managers' skills by Microsoft's TrueSkill algorithm, we find highly skilled managers to behave self-confident resulting in higher risk-taking in the second half of the year compared to less skilled managers. Introducing the TrueSkill algorithm, which is widely used in the e-sports community, to this branch of literature, we can replicate previous findings and theories suggesting overconfidence for mid-years winners.

**Keywords** Fund tournament · Skill · Risk-shifting · TrueSkill · Bayesian network

**JEL Classification** G11 · G23 · G40

## Introduction

Fund managers compete for investor's money by signaling their ability to generate risk-adjusted returns (or alpha) to the market. Using Microsoft's TrueSkill to estimate each manager's skill, we study the impact on the portfolio's risk level. We find highly skilled managers to take systematically more risk within one year's tournament compared to less skilled managers. These results are robust regarding different market phases, different years with pronounced risk-shifting incentives, and different empirical approaches.

Our work contributes to the existing literature by introducing Microsoft's TrueSkill algorithm as a new measure and thus regarding the tournament nature of the fund managers as a "game". Building upon Bayesian network theory, TrueSkill identifies and tracks the skills of managers in a competitive setting in which the belief about a manager's

skill is estimated on the basis of a manager's past performance relative to all other active managers. Despite broad evidence for the long-term underperformance of active managers against a benchmark (Fama 1965), individual managers seem to outperform the market in the short-term, resulting in higher fund inflows and compensation (Sirri and Tufano 1998; Kempf and Ruenzi 2008b), hence promoting a competitive environment among fund managers.

Second, we extend the empirical work in the area of fund tournaments, which was first introduced by Brown et al. (1996). They analyze the behavior of mutual fund managers within one year and detect a risk-seeking investment style for mid-term losers. Replicating their findings, our results indicate winners increasing their risk suggesting a different trend of individual behavior in tournaments in recent years. We then follow Kempf et al. (2009) and highlight risk-shifting differences in years driven by incentives (winners are rewarded for their outperformance) and years driven by unemployment risk (losers are facing high chances of having their funds closed due to underperformance). We extend this area of research by detecting certain investment patterns based on the individual skill level of the managers and highlight the correlation between skill and risk-seeking.

The remainder is structured as follows: In "Fund tournaments and skill" section, we introduce the fund tournaments' setup and Microsoft's TrueSkill, "Empirical results" section contains the empirical analysis, while the final section concludes.

---

✉ Alexander Swade  
a.swade@lancaster.ac.uk

✉ Gerrit Köchling  
gerrit.koechling@udo.edu

Peter N. Posch  
peter.posch@udo.edu

<sup>1</sup> Lancaster University Management School, Lancaster, UK

<sup>2</sup> Invesco Ltd., Frankfurt am Main, Germany

<sup>3</sup> Centre for Finance, Risk and Resource Management, TU Dortmund University, Otto-Hahn-Straße 6, 44227 Dortmund, Germany



## Fund tournaments and skill

### The economics of tournaments

Research in the fields of managerial tournaments is considered as a subset of the agency theoretic contracting theory, which deals with the disparity between principals' and agents' interests and risk aversions. Bolton and Dewatripont (2005) summarize the basic assumptions and implications for multiple scenarios in different areas of economics.

The underlying premise of our analysis is to view the market for portfolio management service as a multi-period decision making. This implies that investors decide in a cyclical pattern which fund service to invest in. One significant aspect of this investment process is the established compensation structure within the fund industry. Fund managers are often compensated based on their funds' assets under management which implies large incentives to generate high fund inflows. Empirical evidence for the positive correlation in a multi-period context of the past performance of the individual fund and new fund inflows has been provided for example by Sirri and Tufano (1998).

This correlation leads to the plain risk adjustment hypothesis in literature of losing managers increasing their risk at mid-term in order to catch up on the leading managers within their peer [cf. Brown et al. (1996)]:

$$\frac{\sigma_{2L}}{\sigma_{1L}} > \frac{\sigma_{2W}}{\sigma_{1W}} \quad (1)$$

where  $\sigma_{pL}$  indicates the risk level of a loser's portfolio in period  $p \in \{1, 2\}$  of a two-period annual tournament and  $\sigma_{pW}$  the risk level of a winner's portfolio, respectively. Multiple researchers followed this hypothesis and analyzed various aspects and implications such as different time periods, competition within fund families, the impact of the selected fund segment, among others. Important ideas and results can be found in the works of Chevalier and Ellison (1997), Busse (2001), Deli (2002), Kempf and Ruenzi (2008a), Kempf and Ruenzi (2008b), and Bär et al. (2010). Despite the findings of all these researchers, there are still contrary opinions about the existence of such tournament behavior between managers and especially the exact behavioral aspects for winners and losers, respectively.

### The impact of prior performance through TrueSkill

New fund inflows are positively correlated with the standings of the individual fund at the end of the tournament, i.e., the end of the year. Most investors tend to trust in the past performance of a fund and expect it to result in positive returns at or around the benchmark level once the fund

claims a top-level within a certain year. Hence, investors update their beliefs about the strength of an individual manager based on past, observed returns, and prior beliefs. In empirical research, this behavior has been modeled for example by Berk and Green (2004), who use a model that includes two key aspects: First, the performance of fund managers is not persistent and, second, investors behave as Bayesians. The first aspect can be interpreted as fund managers are not outperforming a passive benchmark continuously. Second, investors update their belief about the strength of an individual manager based on past, observed returns, and prior beliefs. This leads to the concept of conditional probabilities also known as Bayesian probability where the probability is interpreted as some reasonable expectation based on prior beliefs and knowledge.

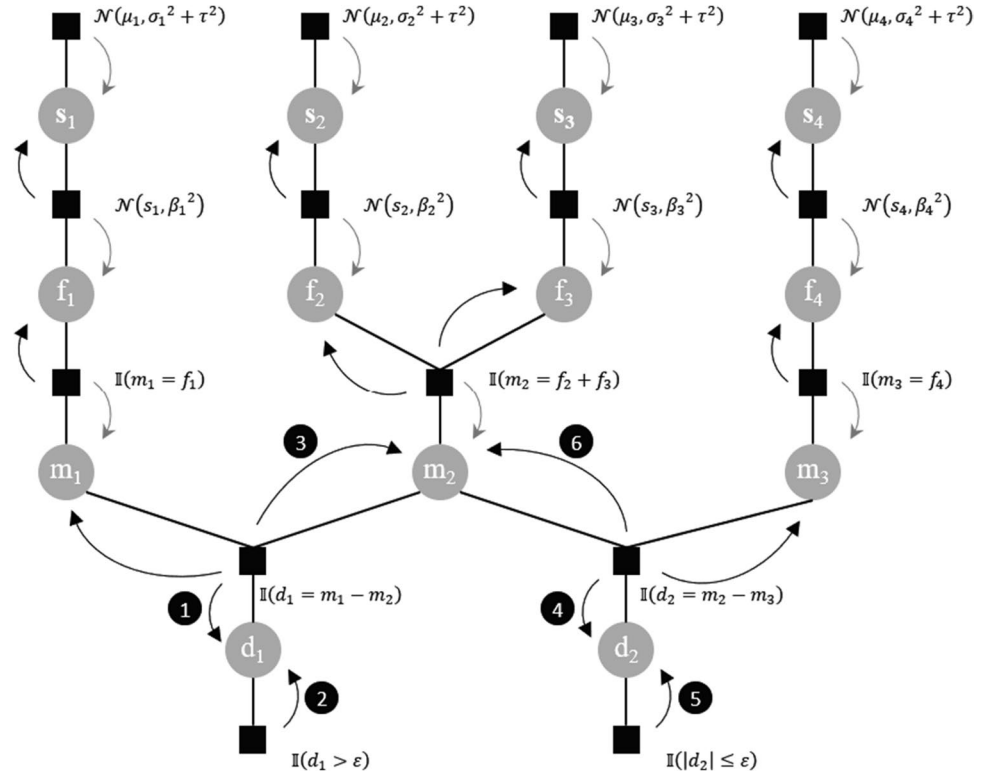
The TrueSkill algorithm has been developed by a team from Microsoft Research in 2005 and is used for match-making in various online games ever since. The purpose of this ranking system is to detect and track the skill of individual players despite playing in teams, derive public rankings, and implement a match-making system that allows players of the same skill to play against each other. The general idea behind TrueSkill is to update the presumption about a player's skill based on the observed outcome of a given game. This technique is called *Bayesian inference* as explained for example by Box and Tiao (2011). TrueSkill characterizes the belief of a manager's skill as Gaussian uniquely described by its mean  $\mu$  and standard deviation  $\sigma$  [cf. Microsoft Research (2005)]. The parameter  $\mu$  can be interpreted as the average manager's skill belief while  $\sigma$  describes the uncertainty about that skill level. The more games a participant plays, the smaller becomes his  $\sigma$  and therefore, the knowledge about a player's skill becomes more precise. Furthermore, his average skill level  $\mu$  is updated based on the match outcome.

One of the most important advantages of TrueSkill is its adaptivity to any underlying setup of ranking match outcomes. It only needs a clear ranking for each match—whether teams are compared with each other or individuals. We will give a brief overview of the underlying process of TrueSkill in order to derive a basic understanding of its functionality. However, we will not explain every mathematical step and its technical realization within the algorithm but refer to the paper of Herbrich et al. (2006).

Let  $k$  managers with a total of  $n$  funds  $\{1, \dots, n\}$  compete in a match. Each fund is uniquely assigned to a single manager resulting in  $k$  disjunct subsets  $A_j \subset \{1, \dots, n\}$ . For each match, the outcome  $\mathbf{r} := (r_1, \dots, r_k) \in \{1, \dots, k\}$  indicates the match specific ranks  $r_j$  for each manager  $j$  in an ascending order; i.e.,  $r_j = 1$  is the winner and possible draws are given as  $r_i = r_j$ . Making use of Bayes' rule, the conditional probability  $P(\mathbf{r}|\mathbf{s}, A)$  of the game outcome  $\mathbf{r}$  given the individual skills  $\mathbf{s} := (s_1, \dots, s_n)$  of all participating funds



**Fig. 1** Schematic work of TrueSkill as a factor graph. Notes: schematic work of TrueSkill illustrated as a factor graph [based on Herbrich et al. (2006)] for the resulting joint distribution  $p(\mathbf{s}, \mathbf{f}, \cdot, \mathbf{r}, A)$  of three managers with a total of four funds and manager 1 winning, while manager 2 and manager 3 draw ( $k = 3$ ,  $A_1 = \{1\}$ ,  $A_2 = \{2, 3\}$ ,  $A_3 = \{4\}$  and the ranking  $\mathbf{r} := (1, 2, 2)$ ). The black boxes represent the factor functions which are used to calculate the local variables—visualized by the light gray circles. The gray arrows indicate the initial calculation of the skill level for all three managers followed by the ‘inner iteration circle.’ This circle is used to approximate the new skill level of all managers, while after that the black arrows indicate the updates of the skill beliefs for each individual fund



in their manager assignments  $A := (A_1, \dots, A_k)$  leads to the posterior distribution of

$$p(\mathbf{s}|\mathbf{r}, A) = \frac{P(\mathbf{r}|\mathbf{s}, A)p(\mathbf{s})}{P(\mathbf{r}|A)}. \quad (2)$$

The prior distribution of the funds' skill  $f(\mathbf{s}) := \prod_{i=1}^n \mathcal{N}(\mu_i, \sigma_i^2)$  is assumed to be a factorizing Gaussian, while each fund  $i$  has a performance  $f_i \sim \mathcal{N}(s_i, \beta_i^2)$  in the match, centered around their individual skill  $s_i$  with fixed variance  $\beta_i^2$ . With TrueSkill, the performance  $m_j$  of a manager  $j$  is defined as the sum of its individual funds' performances indicated by  $m_j := \sum_{i \in A_j} f_i$  [cf. Herbrich et al. (2006)]. Figure 1 shows the exemplary process of TrueSkill as a factor graph. This methodology is used in information technologies to describe complicated 'global' functions consisting of many variables which are most likely derived themselves from various 'local' functions. Those global functions factor as a product of the local functions and can therefore be described in a bipartite graph called factor graph. Further information can be found in Kschischang et al. (2001).

### Identifying skill based tendencies in risk-shifting

In a first step, we calculate the six-month rolling information ratio as a performance measure of each fund. We use these ratios to create a rating of the funds on a monthly base to

feed-forward to the TrueSkill algorithm. At this stage, funds with less than one year of tracking record prior to the start of the tournament year are also included due to the initial calculation of skill levels. Second, the funds included in the annual tournaments compete against each other on a monthly base whereas their skill level—and therefore the skill level of each manager—is calculated by TrueSkill based on the performance rankings. To compare the skill level of different fund managers, we use only each manager's expected average skill level  $\mu$  once the skill development is calculated. To overcome biases for new managers who have not reached their intentional skill level yet, we only consider managers and therefore funds with at least one year of tracking record. This leads to at least 18 matches between all managers and their funds before they are categorized at the end of a tournament's interim period for the first time.

To analyze the skill-dependent risk-shifting, we use conditional transition matrices for the best 20% (high skill), the next 60% (medium skill), and the least 20% (low skill) of each year's managers. We follow the work of Ammann and Verhofen (2009) and adapt this transition approach, commonly known from credit default analyses. The transitions are based on the historical volatility of each manager's portfolio, whereas each manager is assigned to a risk tercile:

$$(e_{i1}, e_{i2}) \in \{1, \dots, 3\}^2, \quad i = 1, \dots, 3 \quad (3)$$



with  $e_{i1}$  characterizing the risk tercile of manager  $i$  in the interim period and  $e_{i2}$  the risk tercile in the second half of the year's tournament. Here, 1 indicates the highest risk tercile and 3 the lowest, respectively. These *migration events* of the same kind are now aggregated in a  $3 \times 3$  matrix of *migration frequencies* where the generic element

$$c_{jk} = \sum_{i=1}^3 \mathbf{1}\{(e_{i1}, e_{i2}) = (j, k)\} \quad (4)$$

is the number of migration events from  $j$  to  $k$  and  $\mathbf{1}\{\dots\}$  the indicator function. Furthermore, we assume that the observations  $e_{i2}$  are the realization of the random variables  $\tilde{e}_{i2}$  with conditional probability distribution

$$p_{jk} = \mathbf{P}(\tilde{e}_{i2} = k \mid \tilde{e}_{i1} = j), \quad \sum_{k=1}^3 p_{jk} = 1 \quad (5)$$

with the probability  $p_{jk}$  of the risk level of a manager's portfolio to change from the  $j$ th to the  $k$ th tercile. Therefore, we use the migration rates as observed:

$$\hat{p}_{jk} = \frac{c_{jk}}{n_j} \quad (6)$$

with  $n_j = \sum_{k=1}^3 c_{jk}$ . To identify any differences between the differently skilled managers, we use a chi-squared test to check for pairwise homogeneity of the transition matrices. The test statistic

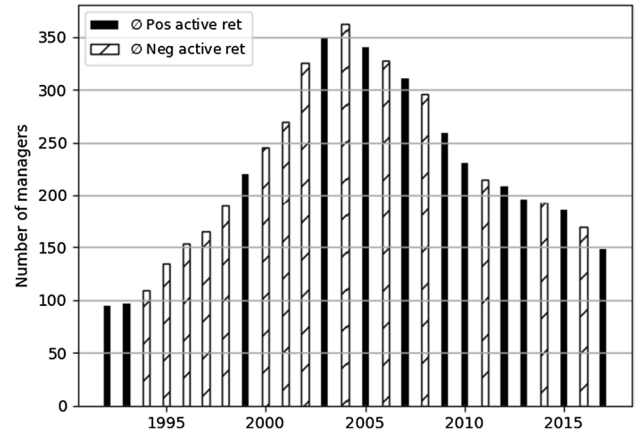
$$\chi^2 = \sum_{k=1}^3 \sum_{s=1}^2 \frac{(c_{jk}(s) - n_j(s)\hat{p}_{jk}^+)^2}{n_j(s)\hat{p}_{jk}^+} \quad (7)$$

is asymptotically  $\chi^2$ -distributed with two degrees of freedom. The variable  $\hat{p}_{jk}^+$  models the estimated probability based on the aggregated data of the two transition matrices and  $s$  is the index for the respective sample, e.g., high- and medium-skilled manager.

By the nature of this approach, our analyses put emphasis on the whole dynamics of the risk-shifting tendencies of differently skilled fund managers. Transition matrices as employed in this study are, inter-alia, widely used in the literature on credit risks [cf. Höse et al. (2009) for details] and in previous studies focusing on prior performance and risk-taking of mutual fund managers (Ammann and Verhoben 2007, 2009).

## Empirical results

Our empirical analysis builds on the two databases Morningstar and Bloomberg. Following Brown et al. (1996) and further researchers in their choice of taking growth-oriented US



**Fig. 2** Managers per Tournament and Benchmark-Related Performance. Notes: black (white) bars indicate a positive (negative) average active return in the respective year

equity funds due to the high interest of financial press and direct investor involvement, we include all funds classified by the Morningstar categories *US OE Large Growth*, *US OE Small Growth* and *US OE Mid-Cap Growth*. We use monthly closing prices by Bloomberg of the categorized funds for the period of 1991 to 2017. This long period allows analyzing the behavior of the managers in various market situations since the selected period combines multiple different aspects such as financial crisis and market phases with a positive long-term trend, e.g., 2009–2017. All funds are listed in US dollar and we clean them for survivorship bias.

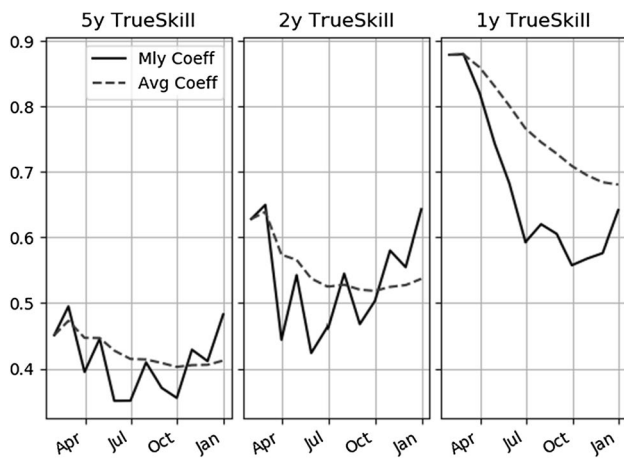
Furthermore, we tackle the fact that various funds are team-managed and multiple managers handle more than one fund by using a string matching algorithm to identify funds managed by the same managers. We exclude all team-managed funds and match the remaining funds clearly to a single manager. This results in 559 individual managers who hold at least one fund on their own within the given time period.

We include all funds in each year's tournament which have at least one year of tracking record and do not miss any data point in the given period. Also, we use two periods of six months to analyze the risk-shifting, which leads to June being the end of the interim period. Those managers above the average at that point are classified as winners and those below as losers. Managers with two or more funds fulfilling these requirements are considered to hold an equally weighted portfolio of their funds to reduce the impact of pro-active risk-shifting across multiple funds. To calculate benchmark-related performance measures, we use the data of the MSCI North America for the same period. An overview of the annual tournaments and the average performance of its participating manager against the benchmark is given in Fig. 2.

There are several options to measure risk-levels of mutual funds. Examples are the return standard deviation,







**Fig. 3** Correlation of TrueSkill and information ratio. Notes: this figure shows the evolution of the Pearson rank correlations between TrueSkill and its underlying performance measure over different time spans for an exemplary year (2015). More precisely, the left (middle, right) figure shows the rank correlation between the TrueSkill rankings estimated over the trailing five (two, one) years and the rankings based solely on the information ratio over the trailing six months to the corresponding month in 2015

two are rather uncommon in mutual fund tournament studies. We follow previous studies and measure risk by the annualized standard deviation of the monthly fund returns (Brown et al. 1996; Kempf and Ruenzi 2008b).

### Measuring performance with TrueSkill

We start our empirical analysis by demonstrating TrueSkill's capability to take prior performance into account. Figure 3 shows the development of the Pearson correlation coefficients between the TrueSkill based rankings of all participating funds within the tournament of five, two, and one years and their information ratio rankings. The left panel shows the correlation with TrueSkill levels being calculated for 4 years prior to 2015, the middle one with 1 year prior, and the right one with TrueSkill establishment just starting in 2015. Hence, Fig. 3 underlines the time dependence of TrueSkill and its adaptation of prior performance while establishing skill levels. Since investors' decisions are often based on behavioral aspects such as prior performance or performance of fund fam-

**Table 1** Risk transitions aggregated 1992–2017

Values in %	Winner			$\chi^2$	Loser			$\chi^2$
Risk tercile	High	Medium	Low		High	Medium	Low	
<i>Panel A: Unconditional</i>								
High	62.4	25.7	11.8		61.3	25.0	13.7	
Medium	30.0	44.8	25.2		26.6	40.9	32.5	
Low	11.1	32.9	56.0		11.2	28.1	60.7	
<i>Panel B: High skill</i>								
High	62.5	27.0	10.5		64.8	23.6	11.5	
Medium	29.0	47.1	24.0		25.8	43.0	31.1	
Low	12.8	31.1	56.2	0.93	12.3	30.4	57.3	2.94
<i>Panel C: Medium skill</i>								
High	63.0	24.6	12.3		61.0	24.8	14.3	
Medium	31.7	42.6	25.6		27.1	42.7	30.2	
Low	11.3	35.0	53.7	2.74	12.2	29.1	58.7	2.08
<i>Panel D: Low skill</i>								
High	60.2	28.0	11.8		59.0	27.1	13.8	
Medium	25.4	49.2	25.4		25.8	34.3	40.0	
Low	8.1	29.1	62.8	7.43**	7.9	24.4	67.8	13.1***

This table shows the risk-shifting tendencies from the year's first-half tercile to the tercile in the second half of the year for the full aggregated dataset. Panel A represents the whole sample, while Panel B to D show the transitions for different skill levels. Each manager is classified as being a winner (loser) if his performance measured by the information ratio lies above (below) the median at the end of the interim period.  $\chi^2$ -values testing  $H_0$ -hypotheses of conditional transitions being equal to the unconditional. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% level (two-tailed tests), respectively

the tracking-error standard deviation which is the standard deviation of the excess returns of the fund over a benchmark, or the systematic risk a fund takes which is commonly estimated via a market model. However, the latter

ily members [e.g., Sirri and Tufano (1998), Nanda et al. (2004)], TrueSkill is an adequate skill measure due to its capability of incorporating these aspects.



## Skill driven risk-shifting

Table 1 shows the aggregated risk-shifting tendencies for the whole sample period. It is structured into four panels—the first one is showing the unconditional transition rates based on the risk terciles in the first and second half of the year and the other three panels are showing the transition rates for the different skill levels. Thus, the three skill-based transition matrices are subsamples of the unconditional case. The  $\chi^2$ -values are representing the  $H_0$ -hypotheses of conditional transitions being equal to the unconditional. Panel D shows significant differences to the unconditional case at the 5% and 1% level for winners and losers, respectively. Indeed, the tendencies in increasing the risk levels are much lower for managers with less skill than for those with high skill.

The first observable pattern is the difference in general risk-seeking between winners and losers in general. While winners tend to stay at their initial level or even increase the risk in the second period of the year, losers act the other way around. With transition rates of 30.0, 11.1, and 32.9 for winner compared to rates of 26.6, 11.2, and 28.1 for losers, Panel A demonstrates the risk-seeking behavior of winners. Vice versa rates of risk decreasing by 25.7, 11.8, and 25.2 compared to 25.0, 13.7, and 32.5 complement this pattern. The subsamples given by Panel B to D indicate similar patterns across the different skill levels. Still, a lot of managers stay within their first half risk tercile with transitions up to 62.4. The transitions for remaining managers are highest for the extreme risk terciles of high and low risk.

Looking at the impact of different skill levels for either tournament standing, we find clear tendencies of high-skilled managers increasing their risk more often than those with less skill regardless of their first-half performance. The comparison of Panel B and Panel D shows higher risk increasing rates for skilled managers in both positions. Hence, the risk decreasing rates are always higher for managers with less skill. Additionally, the risk remaining transitions are bigger for high-skilled managers in the highest risk tercile and low-skilled managers in the lowest risk tercile, respectively.

The subsample for high- and medium-skilled managers are also closely related to the unconditional one. The  $\chi^2$ -test values show no significant differences here. In contrast, the subsample of low-skilled managers differs from the unconditional sample at the 5% level for winners and even at the 1% level for losers. This indicates more controversial behavior for the minority of less-skilled managers, who seem to secure their wins if possible and cut their losses during bad tournaments.

In the next step, we take a closer look at years of extreme risk-shifting. Therefore, we aggregate the five years with the highest risk decreasing by losers and those with the highest risk increasing by losers while winners

acting vice versa. These periods are classified as years dominated by unemployment risk and years dominated by compensation incentives. Hence, we follow Kempf et al. (2009) in their explanation for different risk-shifting tendencies in special periods. The five compensation incentive dominated years are 1992, 1995, 2006, 2014, and 2017, identified in Table 4 in Appendix 1 as those years where the highest RARs are given for mid-year losers. The years dominated by the risk of unemployment are 1993, 2000, 2001, 2004, and 2016; these are the years where losers have extremely low risk adjustment ratios at mid-term.

Table 2 highlights the differences between years dominated by compensation incentives (Panel A) and years dominated by unemployment risk (Panel B). Overall, both panels show similar patterns of winners increasing the risk level in the second half as well. While Panel B has no significant difference between skilled and unskilled winners, Panel A emphasizes the overconfidence of managers with higher skill levels, who are seeking more risk past the interim period. The difference between skilled and unskilled managers is significant at the 10% level.

More importantly, the skill level of individual managers affects their decision making in a losing scenario. Skilled managers seem to rely on their prior performance and increase their risk level dramatically with transitions of 47.8, 8.7, and 34.8 in years dominated by compensation incentives. Instead of cutting their losses, they attempt to catch up by investing very self-confident. Less skilled managers behave differently and cut their losses. They have decreasing transitions of 41.2, 11.8, and 35.7. Here, the skill level of each manager seems to determine his behavior dramatically. The difference between these types of skills is significant at the 1% level. In years dominated by unemployment, the majority of all managers decrease their risk level in the second half of the year, if they lose. Only very few skilled managers try to increase their risk level at this stage and instead, a few managers with less skill start to gamble for a win. Those few seem to go all in before their funds are closed permanently. The difference between skilled and unskilled managers is again significant at the 1% level. Ammann and Verhofen (2007) introduce another dynamic Bayesian network approach to analyze the impact of prior performance. They find similar results given as risk-increasing behavior after years of good performance and decreasing risk-taking after bad years. Within their following work, Ammann and Verhofen (2009) even highlight the same patterns of winning managers increasing their risk in the following period and loser acting vice versa. Our findings are in line with their results and even underlining the impact of prior performance—measured as investor's belief about the individual manager's skill.



**Table 2** Risk transitions based on extreme risk-shifting

Values in%	Winner			Loser		
	High	Medium	Low	High	Medium	Low
<i>Panel A: Compensation incentive dominated</i>						
High skill						
High	57.5	35.0	7.5	72.0	28.0	0.0
Medium	12.5	52.5	35.0	47.8	34.8	17.4
Low	10.0	30.0	60.0	8.7	34.8	56.5
Low skill						
High	43.3	40.0	16.7	47.1	41.2	11.8
Medium	26.9	42.3	30.8	25.0	39.3	35.7
Low	12.5	16.7	70.8	8.9	17.9	73.2
$\chi^2$		5.17*			12.3***	
<i>Panel B: Unemployment risk dominated</i>						
High skill						
High	75.9	24.1	0.0	71.2	15.4	13.5
Medium	31.0	57.1	11.9	4.0	64.0	32.0
Low	0.0	26.4	73.6	0.0	22.2	77.8
Low skill						
High	77.8	18.5	3.7	67.5	25.0	7.5
Medium	25.8	61.3	12.9	18.9	32.4	48.6
Low	4.1	26.5	69.4	2.3	22.7	75.0
$\chi^2$		0.79			9.68***	

This table presents risk transitions for aggregated data of the five years dominated by compensation incentives and unemployment risk, respectively. The selection of the years is based on risk adjustment ratios for each year.  $\chi^2$ -values representing homogeneity test statistic for transitions of high- and low-skilled managers. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% level (two-tailed tests), respectively

## Robustness tests

### Underlying performance measure

The most important parameter within the TrueSkill setup seems to be the choice of the underlying performance measure to calculate monthly rankings, which are the start of further skill calculations. We test for the impact of different performance measures by repeating our analysis with monthly active returns of all participating managers. Table 5 in Appendix 2 shows very similar results to our previous analysis indicating high-skilled managers to increase their risk most of the time regardless of their performance in the first half of the year. These results underline the high correlation between different performance measures as shown by Eling and Schuhmacher (2007). We calculate the Spearman rank-order correlation coefficients inclusive a two-sided  $p$  value for a hypothesis test with the null hypothesis of non-correlation between the data series for three different measures. Table 6 in Appendix 2 outlines the strong and significant correlation between the Sharpe ratio, Information ratio and active return rankings. We conclude that the choice of the underlying performance measure does not affect our initial results significantly.

### Skill thresholds

The results could be driven by the choice of quantiles that classify managers into their skill level. In the main specification, we classified the top 20% as highly skilled and the bottom 20% as low-skilled which leaves a 20–60–20 split. Other reasonable splits, e.g., 10–80–10, lead to the same conclusions as we show in untabulated results.

### Risk adjustment ratio approach

Our next robustness test deals with the general tournament behavior regardless of the individual skill of each manager. Therefore, we replicate the contingency table approach introduced by Brown et al. (1996) based on the risk adjustment ratios. The results presented in Table 4 in Appendix 1 are in line with our results of skill-driven investments, indicating a different trend of individual behavior in tournaments in recent years. Winners have higher RARs in most of the years, which is in contrast to earlier findings of Brown et al. (1996). Still, this demonstrates that our findings are in line with previous methodologies.





## Hyperparameter of the prior distribution

In our empirical analysis, we set the initial prior distribution of the fund managers' skills as described in "The impact of prior performance through TrueSkill" section as  $f(\mathbf{s}) := \prod_{i=1}^n \mathcal{N}(\mu_i, \sigma_i^2)$  with  $\mu_i = 25$  and  $\sigma_i = \frac{\mu_i}{3} \approx 8.33$ . Please note that the average skill level  $\mu_i$  is not of much interest in absolute terms since all managers are assumed to start with the same initial skill. Since we do not define a unit to measure the skill other than using the Gaussian's parameters  $\mu_i$  and  $\sigma_i$ , the relative belief of two fund managers given by their skill distribution is of higher relevance. In that terms, it does not make much difference whether we start with a level of 10, 100, or the standard level of 25<sup>1</sup> as proposed by Herbrich et al. (2006), which originates from TrueSkill's early comparability with the ELO ranking.

To underline the low impact of the initial priors on our results, we vary the relation between  $\mu_i$  and  $\sigma_i$ , i.e.,  $\sigma_i \in \{\frac{\mu_i}{2}, \frac{\mu_i}{4}\}$ . The results are qualitatively similar to our base case  $\sigma_i = \frac{\mu_i}{3}$ , see Tables 7 and 8 in Appendix 2.

The neglectable impact of the priors is in line with the theoretical expectation about their impact: With sample size  $n \rightarrow \infty$  the difference between two posteriors based on different Gaussian priors tends toward zero. The same holds for larger prior variances  $\sigma_i$ , as outlined for example by Ley et al. (2017).

## Different benchmark indexes

Within our analysis, we use a risk-adjusted approach to determine the rankings of each manager used for the TrueSkill algorithm. In fact, our measure of choice is the information ratio as a market model adjustment measure where the benchmark is the MSCI North America. Given the different setup of mutual funds and their long-term purposes, e.g., equity-only, long-only, multi-asset, and so on, our chosen benchmark might not be appropriate for every mutual fund in the universe. Nevertheless, we restrict our fund sample to growth-oriented US equity mutual funds as earlier researchers before (Brown et al. 1996; Taylor 2003; Kempf and Ruenzi 2008b). The categorization is based on the widely accepted classification by Morningstar, which leads to a quite homogeneous sample. We qualify this putative sample restriction by similar arguments used in earlier research.

However, Morningstar specifies two benchmark indexes for each of its categories. The primary index for all three

categories used in this study is the S&P500 which correlates almost perfectly with the MSCI North America. The secondary benchmark index differs for each category.<sup>2</sup> We repeat our analysis benchmarking each fund on its secondary benchmark index and report the results in Table 9. Overall, the conditional transition matrices differ stronger from the unconditional transition matrix than in our baseline case. In line with our previous findings, we find a tendency that winning managers increase their risk more than losers and that managers classified as low-skilled seem to adjust their risk less than managers classified as high-skilled.

## Regression approach

On the basis of the conditional transition matrix approach, our results suggest that the risk-shifting tendencies are significantly different for low- and high-skilled fund managers and, beyond that, that high-skilled managers tend to increase their risk-levels to a higher extent compared to low-skilled managers. We acknowledge that conclusions like these have to be interpreted with caution due to unobservable covariates that might influence the results. To mitigate the effect of omitted variables and provide further empirical evidence for our conclusions, we formulate the following regression model:

$$\Delta\sigma_{i,t} = \beta_1 \text{Rank}_{i,t} \times D_{i,t}^H + \beta_2 \text{Rank}_{i,t} \times D_{i,t}^L + \beta_3 \sigma_{i,t}^{\text{First Half}} + \epsilon_{i,t} \quad (8)$$

where the dependent variable,  $\Delta\sigma_{i,t} = \sigma_{i,t}^{\text{Second Half}} - \sigma_{i,t}^{\text{First Half}}$ , is the change in standard deviations of fund  $i$ 's returns from the first to the second half of the year  $t$ .  $\text{Rank}_{i,t}$  denotes the rank of the fund manager with respect to all other managers scaled to the interval  $[0, 1]$  (1 being best). High respectively low manager skill is denoted by  $D_{i,t}^*$  with  $* \in \{H, L\}$ . In a further specification, we replace  $\text{Rank}_{i,t}$  with dummy variables indicating that a fund manager ranked in the top 20% respectively bottom 20% of all active managers analogous to the main analysis. For all specifications, we include time and fund-company fixed effects. The latter control, for example, for all time-invariant characteristics attributable to a manager's company that may influence the results.

We present the results of four specifications in Table 3, two each using either the information ratio or active returns to estimate the managers' skill levels via TrueSkill. All specifications indicate that high-skilled fund managers significantly increase their risk after performing well in the first half of the year. Contrary, we find the opposite signs for any coefficient associated with risk-shifting of less-skilled fund managers. Equality tests reject the null hypotheses  $D^{\text{Win}} \times D^H = D^{\text{Win}} \times D^L$  and  $D^{\text{Loss}} \times D^H = D^{\text{Loss}} \times D^L$ . The explained variation in risk-shifting amounts to  $\approx 75\%$ , which is a common value in fund tournament studies. Overall, the results support our conclusions drawn from the conditional

<sup>1</sup> We confirm this assumption in unreported results.

<sup>2</sup> Secondary benchmark indexes are Russell 1000 Growth, Russell Mid-Cap Growth, and Russell 2000 Growth for the Large Growth, Mid-Cap Growth, respectively, Small Growth category.



**Table 3** Regression models of different skill-levels on risk-shifting

	Information ratio		Active return	
	(1)	(2)	(3)	(4)
$\text{Rank} \times D^H$	0.005*		0.006*	
	(1.889)		(1.940)	
$\text{Rank} \times D^L$	− 0.003		− 0.003	
	(− 0.584)		(− 0.689)	
$D^{\text{Win}} \times D^H$		0.003*		0.005**
		(1.747)		(2.060)
$D^{\text{Win}} \times D^L$		− 0.003		− 0.003
		(− 1.042)		(− 0.986)
$D^{\text{Loss}} \times D^H$		0.003		0.003
		(0.909)		(0.917)
$D^{\text{Loss}} \times D^L$		− 0.004		− 0.008***
		(− 1.580)		(− 2.854)
$\sigma_{\text{First Half}}$	− 0.361***	− 0.363***	− 0.362***	− 0.365***
	(− 5.116)	(− 5.124)	(− 5.114)	(− 5.143)
Fixed effects	Yes	Yes	Yes	Yes
Observations	5155	5155	5157	5157
Adjusted R <sup>2</sup>	0.747	0.747	0.747	0.748

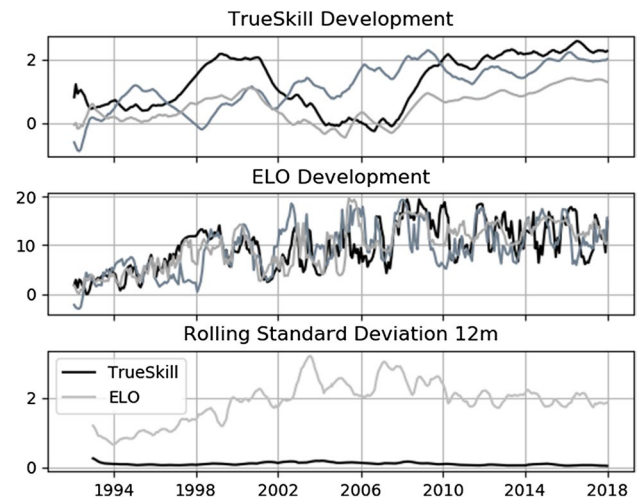
This table presents results of a regression of fund managers' performance in the first half of the year on their risk-shifting in the second half, conditional on their estimated skill levels (high, low). Rank denotes the rank of the fund manager scaled to the interval [0, 1] (1 being best) and D indicate dummy variables for high or low skill, or for ranking among the top 20% best or worst managers. Robust standard errors are clustered by year. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% level (two-tailed tests), respectively

transition matrix approach and provide further insights on the channels that foster the results.

dummy

### Comparison to ELO

Last, we compare TrueSkill with another popular skill measure—the ELO rating, most known from the world of chess. The ELO ranking system is used in competitive chess as well as various unofficial rankings, e.g., online gaming or football tournaments. It is much simpler in its calculations and therefore not capable to adapt teams playing each other. Figure 4 shows the skill development of three random managers of the whole period sample measured by TrueSkill and ELO. Both ratings are based on monthly matches between all managers participating in a given year's tournament. The skill levels are normalized to make them comparable since the absolute level differs between both systems. Due to our premise of being included in a year's tournament if and only if there is more than one year of tracking record, the managers seem to start with different levels, but in fact, they started all with the same setup, initially. The ELO ratings vary rapidly on a high frequency, while the TrueSkill



**Fig. 4** Different skill development of three random fund managers. Notes: this figure shows the temporal development of skill ratings based on TrueSkill and ELO for three randomly chosen fund managers from our sample. ELO is a method for calculating the relative skill levels, commonly used in chess. The bottom figure compares the rolling standard deviations of the two methods highlighting the stable skill belief estimated via TrueSkill

ratings are adjusting themselves much slower and only react to unexpected outcomes.

The third panel of Fig. 4 underlines the differences in volatility by representing the rolling standard deviation over 12 months of the normalized ELO and TrueSkill ratings, respectively. Hence, the average TrueSkill standard deviation is at 0.106 and therefore much lower than the one of ELO given as 1.907. A good skill measure should offer low volatility to establish a stable belief about the skill level of an individual manager in the long-term. The time stability of TrueSkill shows its potential to classify managers into skill levels and derive skill-based behavior from it.

Summarizing the results of the robustness checks, our results about the impact of skill are in line with theories in behavioral finance and psychology, showing the overconfidence of outperforming managers in their investment decisions. Taylor and Brown (1988) find evidence of people having unrealistically positive views of themselves which leads to the described self-confidence not only after being among the winners for a couple of competitions and De Bondt and Thaler (1995) detect a positive correlation between high confidence and above average-trading frequencies.

### Conclusion

Our results highlight the self-confident behavior of skilled managers by holding or increasing their portfolio risk in almost every situation compared to those with less skill. Applying the TrueSkill algorithm to display



**Table 4** Contingency tables of annual tournaments

<i>Freq. in%</i>		Winner		Loser		$\chi^2$ – value <sup>b</sup>	<i>p</i> value
Year	No. <sup>a</sup>	High RAR	Low RAR	High RAR	Low RAR		
1992	94	19.2	30.9	30.9	19.2	5.1489	0.1612
1993	97	33.0	17.5	17.5	32.0	9.6907	0.0337
1994	109	22.9	27.5	27.5	22.0	1.1284	0.7702
1995	135	20.2	30.6	29.9	19.4	5.8806	0.1176
1996	153	25.8	25.2	24.5	24.5	0.0728	0.9949
1997	165	27.9	22.4	22.4	27.3	1.7636	0.6229
1998	189	26.1	24.5	23.9	25.5	0.2128	0.9755
1999	220	25.0	25.0	25.0	25.0	0.0000	1.0000
2000	245	33.1	17.1	17.1	32.7	24.208	0.0000
2001	269	34.2	16.0	16.0	33.8	34.985	0.0000
2002	325	25.9	24.3	24.3	25.5	0.2552	0.4625
2003	350	27.1	22.9	22.9	27.1	2.5714	0.4625
2004	362	30.4	19.6	19.6	30.4	16.807	0.0008
2005	340	23.0	26.8	27.1	23.0	2.1563	0.5406
2006	328	20.2	30.0	30.0	19.9	12.927	0.0048
2007	311	30.4	20.1	19.7	29.8	12.877	0.0049
2008	296	27.8	22.0	22.4	27.8	3.6983	0.2959
2009	259	24.9	25.7	25.3	24.1	0.1362	0.9872
2010	231	26.0	24.2	24.2	25.5	0.2208	0.9742
2011	215	27.0	23.3	23.3	26.5	1.0558	0.7878
2012	208	27.4	22.6	22.6	27.4	1.9231	0.5885
2013	196	24.5	25.5	25.5	24.5	0.0816	0.9930
2014	193	21.2	29.0	29.0	20.7	4.9896	0.1726
2015	185	31.3	18.9	18.9	30.8	10.957	0.0120
2016	169	33.1	17.2	17.2	32.5	16.633	0.0008
2017	148	21.0	29.1	29.1	21.0	3.8919	0.2734

<sup>a</sup> Number of managers within the given tournament year<sup>b</sup> Based on the null hypothesis of every cell receiving 25% of the distribution

Notes: This table presents the annual contingency tables of risk adjustment ratios (RAR) for the whole dataset covering the years 1992–2017. Managers who perform at the end of the interim period above (below) the median are classified as winners (losers). The same methodology applies to the classification of high (low) RARs

investors' beliefs about the individual skill level of fund managers, we present a way to model the positive correlation of prior performance and new investment decisions.

The impact of good performance in recent years seems to lead to an over-confident investment style of managers, who are shifting their portfolio risk toward the higher tercile of the peer group in the second half of the year. Only a few managers classified with less skill increase their risk in a losing situation. We demonstrate the robustness of our results regarding the choice of the performance measure to rank the managers each month as well as the usability of TrueSkill as an adequate representation of investor's belief about a manager's skill.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



## Appendix 1: General tournament behavior

To detect general tournament behavior, we follow Brown et al. (1996) by using their contingency table approach to determine the risk adjustments during the second half of the tournament: The performance of every manager  $i$  is given as the information ratio  $IR_{iM}$  against the MSCI North America in the first 6 months of the year to identify mid-term winners and losers as those above or below the median information ratio, respectively. All managers hold an equally weighted portfolio of their funds  $j \in \{1, \dots, n\}$  included in the tournament with a portfolio return of  $r_{ik}^{\text{port}}$  (Table 4).

In order to calculate the information ratio for each fund given as  $IR = \frac{\text{active premia}}{\text{tracking error}}$ , we determine the cumulative return as  $RTN_i = \prod_{k=1}^6 (1 + r_{ik}^{\text{port}}) - 1$ , which finally leads to the formula for the information ratio of manager  $i$  at the end of June

$$IR_i = \frac{RTN_i - RTN_b}{\sqrt{\frac{1}{6-1} \sum_{k=1}^6 (r_{ik}^{\text{port}} - r_{bk})^2}} \quad (9)$$

The variable  $RTN_b$  is the cumulative return of benchmark returns  $r_{bk}$  of month  $k \in \{1, \dots, 6\}$ . The risk

adjustment ratio (RAR) of manager  $i$  for the given tournament year with interim assessment date in June is:

$$RAR_i = \sqrt{\frac{\sum_{k=6+1}^{12} (r_{ik}^{\text{port}} - \bar{r}_{i(12-6)})^2}{(12-6)-1}} \div \sqrt{\frac{\sum_{k=1}^6 (r_{ik}^{\text{port}} - \bar{r}_i)^2}{6-1}} \quad (10)$$

with  $\bar{r}_{i(12-6)}$  and  $\bar{r}_i$  representing the mean portfolio return of fund manager  $i$  before and after the assessment date, respectively. This variable measures the risk adjustments of a given portfolio within the two periods of the year's tournament by comparing the portfolio's volatility in both periods. Thus, we rank the RAR in a similar way as the IR and determine high RAR as those above the median and low RAR as those below the median for the first and second period, respectively.

## Appendix 2: Tables—robustness tests

See Tables 5, 6, 7, 8 and 9.

**Table 5** Risk transitions aggregated 1992–2017—active return

Values in %	Winner			$\chi^2$	Loser			$\chi^2$
Risk tercile	High	Medium	Low		High	Medium	Low	
<i>Panel A: Unconditional</i>								
High	62.6	25.9	11.5		61.1	24.9	14.0	
Medium	31.0	44.7	24.2		25.5	40.9	33.7	
Low	12.1	32.2	54.8		10.2	27.9	61.9	
<i>Panel B: High skill</i>								
High	64.1	26.4	9.5		66.5	24.1	9.4	
Medium	35.5	45.0	19.5		26.9	43.8	29.4	
Low	16.2	34.7	49.1	5.96*	14.2	29.1	56.7	7.18**
<i>Panel C: Medium skill</i>								
High	61.7	25.2	13.1		59.3	25.3	15.4	
Medium	29.4	44.4	26.2		26.1	41.6	32.3	
Low	11.1	32.2	56.7	1.92	10.3	29.4	60.3	2.06
<i>Panel D: Low skill</i>								
High	63.5	27.6	8.8		61.0	24.5	14.5	
Medium	31.0	45.7	23.4		22.2	36.1	41.7	
Low	9.7	34.8	55.5	0.64	7.3	22.8	69.9	13.63***

This table shows the risk-shifting tendencies from the year's first-half tercile to the tercile in the second half of the year for the full aggregated data set. Panel A represents the whole sample, while Panel B to D show the transitions for different skill levels. Each manager is classified as being a winner (loser) if his performance measured by the active return lies above (below) the median at the end of the interim period.  $\chi^2$ -values testing H0-hypotheses of conditional transitions being equal to the unconditional. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% level (two-tailed tests), respectively



**Table 6** Rank correlations based on different performance measures

Performance measures	Sharpe ratio	Active return	Information ratio
Sharpe ratio		1	
Active return	0.983***	1	
Information ratio	0.943***	0.963***	1

The ranks are calculated over a time period from 1992 to 2017. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% level (two-tailed tests), respectively

**Table 7** Results for different hyperparameters:  $\sigma = 12.5$ 

Values in%	Winner			$\chi^2$	Loser			$\chi^2$
Risk tercile	High	Medium	Low		High	Medium	Low	
<i>Panel A: Unconditional</i>								
High	62.4	25.7	11.8		61.3	25.0	13.7	
Med	30.0	44.8	25.2		26.6	40.9	32.5	
Low	11.1	32.9	56.0		11.1	28.2	60.7	
<i>Panel B: High skill</i>								
High	62.6	26.7	10.8		65.8	23.8	10.4	
Med	29.2	46.6	24.2		26.8	42.5	30.7	
Low	13.0	31.3	55.7	6.35**	13.5	31.8	54.7	0.66
<i>Panel C: Medium skill</i>								
High	62.7	25.0	12.3		61.0	25.4	13.6	
Med	32.1	42.7	25.2		27.2	42.9	29.9	
Low	11.5	34.7	53.7	2.75	12.2	29.4	58.3	2.63
<i>Panel D: Low skill</i>								
High	61.1	27.4	11.5		57.6	25.3	17.2	
Med	23.9	49.4	26.7		24.8	34.8	40.5	
Low	7.3	29.6	63.1	22.25***	6.8	22.5	70.8	8.58**

This table shows the risk-shifting tendencies from the year's first-half tercile to the tercile in the second half of the year for the full aggregated data set for a different choice of hyperparameter for the prior distribution of the fund managers' skills, e.g.,  $\sigma = 12.5$ . Panel A represents the whole sample, while Panel B to D show the transitions for different skill levels. Each manager is classified as being a winner (loser) if his performance measured by the active return lies above (below) the median at the end of the interim period.  $\chi^2$ -values testing H0-hypotheses of conditional transitions being equal to the unconditional. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% level (two-tailed tests), respectively





**Table 8** Results for different hyperparameters:  $\sigma = 6.25$ 

Values in%	Winner			$\chi^2$	Loser			$\chi^2$
Risk tercile	High	Medium	Low		High	Medium	Low	
<i>Panel A: Unconditional</i>								
High	62.4	25.7	11.8		61.3	25.0	13.7	
Med	30.0	44.8	25.2		26.6	40.9	32.5	
Low	11.1	32.9	56.0		11.1	28.2	60.7	
<i>Panel B: High skill</i>								
High	64.7	23.9	11.4		64.1	22.2	13.8	
Med	26.9	46.2	26.9		23.7	41.0	35.3	
Low	11.2	36.2	52.6	4.36	11.8	32.4	55.9	3.42
<i>Panel C: Medium skill</i>								
High	61.8	26.3	11.9		60.8	25.9	13.3	
Med	33.1	43.1	23.8		27.0	43.6	29.4	
Low	12.0	32.6	55.3	3.50	11.9	29.6	58.5	1.98
<i>Panel D: Low skill</i>								
High	61.8	26.1	12.1		60.5	24.9	14.6	
Med	23.9	48.3	27.8		27.7	33.0	39.4	
Low	8.0	29.3	62.6	16.75***	9.0	22.3	68.8	7.26**

Notes: This table shows the risk-shifting tendencies from the year's first-half tercile to the tercile in the second half of the year for the full aggregated data set for a different choice of hyperparameter for the prior distribution of the fund managers' skills, e.g.,  $\sigma = 6.25$ . Panel A represents the whole sample, while Panel B to D show the transitions for different skill levels. Each manager is classified as being a winner (loser) if his performance measured by the active return lies above (below) the median at the end of the interim period.  $\chi^2$ -values testing H0-hypotheses of conditional transitions being equal to the unconditional. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% level (two-tailed tests), respectively

**Table 9** Risk transitions aggregated 1992–2017—secondary benchmarks

Values in%	Winner			$\chi^2$	Loser			$\chi^2$
Risk tercile	High	Medium	Low		High	Medium	Low	
<i>Panel A: Unconditional</i>								
High	62.4	25.7	11.8		61.3	25.0	13.7	
Med	30.0	44.8	25.2		26.6	40.9	32.5	
Low	11.1	32.9	56.0		11.1	28.2	60.7	
<i>Panel B: High skill</i>								
High	68.7	23.6	7.7		67.5	23.4	9.1	
Med	31.6	43.1	25.4		24.2	38.8	37.1	
Low	16.5	34.6	48.9	8.65**	12.8	31.5	55.7	10.52***
<i>Panel C: Medium skill</i>								
High	60.3	25.7	14.0		57.8	27.1	15.1	
Med	29.2	44.7	26.1		28.2	43.1	28.7	
Low	10.8	33.0	56.1	5.78*	10.4	29.2	60.4	1.35
<i>Panel D: Low skill</i>								
High	60.9	28.7	10.3		65.0	20.7	14.3	
Med	30.7	46.5	22.8		24.0	36.3	39.8	
Low	6.5	30.8	62.7	9.90***	12.3	22.6	65.1	5.61*

This table shows the risk-shifting tendencies from the year's first-half tercile to the tercile in the second half of the year for the full aggregated data set for a different choice of benchmarks for the funds in the sample. While in the baseline specification all funds are benchmarked against the MSCI North America, they are now benchmarked against different Russel indexes, see "Empirical results" section. Panel A represents the whole sample, while Panel B to D show the transitions for different skill levels. Each manager is classified as being a winner (loser) if his performance measured by the active return lies above (below) the median at the end of the interim period.  $\chi^2$ -values testing H0-hypotheses of conditional transitions being equal to the unconditional. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% level (two-tailed tests), respectively



## References

- Ammann, M., and M. Verhofen. 2007. Prior performance and risk-taking of mutual fund managers: A dynamic bayesian network approach. *Journal of Behavioral Finance* 8 (1): 20–34.
- Ammann, M., and M. Verhofen. 2009. The impact of prior performance on the risk-taking of mutual fund managers. *Annals of Finance* 5 (1): 69–90.
- Bär, M., A. Kempf, and S. Ruenzi. 2010. Is a team different from the sum of its parts? Evidence from mutual fund managers. *Review of Finance* 15 (2): 359–396.
- Berk, J.B., and R.C. Green. 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112 (6): 1269–1295.
- Bolton, P., M. Dewatripont, et al. 2005. *Contract theory*.
- Box, G.E., and G.C. Tiao. 2011. *Bayesian inference in statistical analysis*, vol. 40. Hoboken: Wiley.
- Brown, K.C., W. Harlow, L.T. Starks, et al. 1996. Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. *Journal of Finance* 51 (1): 85–110.
- Busse, J.A. 2001. Another look at mutual fund tournaments. *Journal of Financial and Quantitative Analysis* 36 (1): 53–73.
- Chevalier, J., and G. Ellison. 1997. Risk taking by mutual funds as a response to incentives. *Journal of Political Economy* 105 (6): 1167–1200.
- De Bondt, W.F., and R.H. Thaler. 1995. Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in Operations Research and Management Science* 9: 385–410.
- Deli, D.N. 2002. Mutual fund advisory contracts: An empirical investigation. *Journal of Finance* 57: 109–133.
- Eling, M., and F. Schuhmacher. 2007. Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking & Finance* 31 (9): 2632–2647.
- Fama, E.F. 1965. The behavior of stock-market prices. *Journal of Business* 38 (1): 34–105.
- Herbrich, R., T. Minka, and T. Graepel 2006. TrueSkill™: A Bayesian skill rating system. In *Proceedings of the 19th international conference on neural information processing systems*, 569–576. MIT Press.
- Höse, S., S. Huschens, and R. Wania. 2009. Rating migrations. In *Applied quantitative finance*, 105–123. Springer.
- Kempf, A., and S. Ruenzi. 2008a. Family matters: rankings within fund families and fund inflows. *Journal of Business Finance & Accounting* 35 (1–2): 177–199.
- Kempf, A., and S. Ruenzi. 2008b. Tournaments in mutual-fund families. *The Review of Financial Studies* 21 (2): 1013–1036.
- Kempf, A., S. Ruenzi, and T. Thiele. 2009. Employment risk, compensation incentives, and managerial risk taking: Evidence from the mutual fund industry. *Journal of Financial Economics* 92 (1): 92–108.
- Kschischang, F.R., B.J. Frey, and H.-A. Loeliger. 2001. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory* 47 (2): 498–519.
- Ley, C., G. Reinert, Y. Swan, et al. 2017. Distances between nested densities and a measure of the impact of the prior in Bayesian statistics. *The Annals of Applied Probability* 27 (1): 216–241.
- Microsoft Research. 2005. TrueSkill ranking system.
- Nanda, V., Z.J. Wang, and L. Zheng. 2004. Family values and the star phenomenon: Strategies of mutual fund families. *Review of Financial Studies* 17 (3): 667–698.
- Sirri, E.R., and P. Tufano. 1998. Costly search and mutual fund flows. *Journal of Finance* 53 (5): 1589–1622.
- Taylor, J. 2003. Risk-taking behavior in mutual fund tournaments. *Journal of Economic Behavior & Organization* 50 (3): 373–383.
- Taylor, S.E., and J.D. Brown. 1988. Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin* 103 (2): 193.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Alexander Swade** graduated in Business Mathematics (M.Sc.) with a specialization in Quantitative Finance at the TU Dortmund University in 2019. During his studies, he got further insights into multi-asset factor investing as a working student for Bankhaus Lampe KG in Düsseldorf, Germany. Currently, he is pursuing a Ph.D. in Finance at Lancaster University. In a joint research initiative between Lancaster University and Invesco Quantitative Strategies, Alexander investigates factor-based investment strategies across different asset classes and within equities.

**Gerrit Köchling** is a research associate at Centre for Finance, Risk and Resource Management and a third-year Finance Ph.D. student at the technical university of Dortmund. He graduated in Business Mathematics in 2017 with the main focus on mathematical optimization and finance, and served prior to that as an intern at the KfW banking group in Frankfurt am Main, Germany. His research covers a wide area in finance especially focussing on corporate and behavioral finance, as well as FinTech. Gerrit is passionate about a wide range of data-driven approaches with a special interest in text mining.

**Peter Nikolaus Posch** is professor of finance at the technical university of Dortmund, Germany, and director of the Centre for Finance, Risk and Resource Management (FiRRM). Previously he served as co-Head of credit treasury at a top-5 bank where he traded for the proprietary book and was responsible for the active management of the group's credit risk. He studied quantitative economics, philosophy, and law at the University of Bonn. Peter is a frequent speaker at international conferences, a trainer for several companies, and conducts consultancy projects for financial institutions and corporates.

