CFR Working Paper NO. 19-04

Drawdown Measures: Are They All the Same?

O. Korn • P. Moeller • Ch. Schwehm

centre for financial research
cologne

Drawdown Measures: Are They All the Same?*

Olaf Korn[†], Philipp M. Möller[‡], Christian Schwehm[§]

Current Version: October 2019

JEL Classification: G11

Keywords: Asset Management, Drawdown, Risk Measures, Performance Measurement

*We would like to thank Antti Ilmanen and Jan Viebig for helpful comments and suggestions.

[†]Olaf Korn, University of Goettingen and Centre for Financial Research Cologne (CFR), Platz der Göttinger Sieben 3, D-37073 Göttingen, Germany, Phone +49 551 39 7262, Fax +49 551 39 7665, Email okorn@uni-goettingen.de

[‡]Philipp M. Möller, University of Goettingen, Platz der Göttinger Sieben 3, D-37073 Göttingen, Germany, Phone +49 551 39 21169, Fax +49 551 39 7665, Email philipp.moeller@uni-goettingen.de

[§]Christian Schwehm, Quoniam Asset Management GmbH, Westhafenplatz 1, D-60327 Frankfurt am Main, Germany, Phone +49 69 74384 267, Fax +49 69 74384 222, Email christian.schwehm@quoniam.com

Abstract

Over the years, a diverse range of drawdown measures has evolved to guide asset management. We show that almost all of these measures fit into a unified framework. This new framework simplifies the implementation of drawdown measures and improves understanding their similarities and differences. Conceptual differences between drawdown measures translate into different rankings of portfolios, which we document in a simulation study. Our research also shows that all drawdown measures can (to some degree) discriminate between skillful and unskillful portfolio managers, but differ in terms of accuracy. However, the ability to detect skill does not easily improve performance ratios where drawdown measures serve as the denominator. In conclusion, our study shows that the choice of an adequate drawdown measure is vital to the assessment of investments because different measures emphasize different aspects of risk.

I Introduction

Drawdown measures quantify risk by penalizing losses from previous gains. They capture important aspects of what investors consider 'risk', including psychological aspects (e.g., regret), which are central to financial decision making (Frydman and Camerer, 2016). Drawdown measures are path-dependent by construction, which sets them apart from other risk measures, such as (semi-)variance, value-at-risk, or expected shortfall. Therefore, drawdown measures complement classical risk measures in important ways and are widely used in asset management. Driven by industry[1] and academia, a wide variety of drawdown measures has been developed, including maximum drawdown, average drawdown, conditional drawdown, conditional expected drawdown, average squared drawdown, and end-of-period drawdown. (Martin and McCann, 1989; Chekhlov et al., 2005; Goldberg and Mahmoud, 2017; Möller, 2018). How similar are these measures? Do they all lead to the same conclusions? If these measures do differ, what should guide a specific investor's choice of an appropriate drawdown measure in a specific situation? Our paper provides answers to these questions both from a theoretical and empirical angle.

As a theoretical contribution, we establish that almost all drawdown measures can be subsumed under a common framework, which we refer to as the *weighted drawdown* (wDD) framework because its main idea is to attach weights to different elements of the drawdown graph. We explicitly show how to choose these weights to recover various drawdown measures. The weights themselves provide information about the economic idea behind each measure. Additionally, a comparison of weights offers a straightforward way of discovering differences and similarities between drawdown measures. The wDD framework is also useful in the implementation of drawdown measures because a generic computer code can simply be adapted to alternative weight functions to obtain different measures. A further benefit of the wDD framework is that it not only enhances the understanding of existing drawdown measures, but also provides an easy tool to construct customized drawdown measures. By choosing a set of weights, new drawdown measures can be developed and tailored to a client's conception of risk.

[1] The drawdown concept was first floated and propagated by finance practitioners, such as Young (1991), Burke (1994), Kestner (1996), Odo (2006), or Schmielewski and Schwehm (2014), in a quest to find risk measures that are relevant to investors.

In our empirical study, we quantify the degrees of similarity between various drawdown measures. Using almost 20 years of MSCI World index data, we simulate the behavior of portfolio managers who assemble stock portfolios under various realistic constraints. For the resulting portfolio strategies, we then compute rank correlations, that is, we compare how each drawdown measure ranks these portfolio strategies in comparison to other drawdown measures.[2] Consistent with the intuition from the wDD framework, our empirical results reveal a nuanced system of relationships between the different drawdown measures. Notably, the results show that average drawdown, average squared drawdown, and linearly weighted drawdown are closely related, but that correlations drop significantly when it comes to maximum drawdown and end-of-period drawdown. Hence, different drawdown measures potentially yield substantially different rankings of investments and are not all the same.

In the base setting of our empirical study, all portfolio managers pick stocks from the index purely at random. Additionally, we model skillful and unskillful managers by assigning different hit ratios—i.e., probabilities to pick future winners.[3] The analysis of management skill is important because a lack of skill represents a crucial aspect of risk that drawdown measures should be able to detect. Our empirical results show that all drawdown measures are indeed useful in skill discrimination. They are capable of detecting skill because they capture more aspects about risk than variability alone. However, while average drawdown and linearly weighted drawdown are particularly useful, maximum drawdown and end-of-period drawdown are considerably weaker at differentiating between skillful and unskillful managers.

It is a natural idea to exploit the ability of drawdown measures to detect skill for the improvement of performance ratios.[4] The question as to whether ratios, which use different drawdown measures in the denominator, truly differ from each other has been asked repeatedly in the

[2]Rank correlations are a common method for assessing similarities (see, for example, Eling (2008), Haas Ornelas et al. (2012), or Auer (2015)).

[3]Since Jensen (1968), a vast body of literature has studied manager skill using different empirical techniques (Grinblatt and Titman, 1989; Fama and French, 2010; Berk and van Binsbergen, 2015). By varying the hit ratio, our simulation model incorporates skill in a very intuitive way without requiring complex assumptions about the data generating process.

[4]In performance measurement, it is a common technique to divide excess returns by measures of risk to obtain performance ratios (e.g., Caporin et al. (2014)).

literature.[5] Our empirical study based on portfolio rankings via performance ratios shows that differences exist between drawdown measures, although rankings via performance ratios are more similar than rankings via risk measures.[6] Surprisingly, drawdown performance measures do not improve the detection of skill as compared to the Sharpe ratio. This is due to deficiencies in performance ratios in general and is not specific to drawdown measures. To the contrary, the ability of drawdown measures to detect skill becomes a drawback in the ranking of portfolios if observed returns become negative. Therefore, a naïve application of drawdown performance measures is not recommended.

Given all the empirical evidence, the question as to whether drawdown measures are all the same can be answered in the negative, which is in line with intuition from the wDD framework. While all drawdown measures produce portfolio rankings that are positively correlated, maximum drawdown and end-of-period drawdown display significantly different results. All drawdown measures can be used to differentiate between skillful and unskillful managers, but average drawdown and linearly weighted drawdown outperform their peers. Differences in portfolio rankings and skill detection also appear in performance ratios; however, results based on drawdown performance ratios are more similar than the results based on the drawdown measures themselves.

II A Unified Framework of Drawdown Measures

To date, many different drawdown-based risk measures have been introduced. The widely used maximum drawdown (MDD) has been applied in portfolio management at least since the 1980s and measures the single largest peak to trough loss (Garcia and Gould, 1987). Like the average drawdown (ADD), it belongs to the conditional drawdown family introduced by Chekhlov et al. (2005). The average squared drawdown (ADD^2), also called Ulcer index, was introduced by

[5]Eling and Schuhmacher (2007), Eling (2008), Caporin and Lisi (2011), Haas Ornelas et al. (2012), Auer and Schuhmacher (2013), and Auer (2015) conclude that these drawdown ratios are essentially the same because they lead to the same rankings of investments.

[6]The conclusions drawn by this study may differ from those drawn by previous research, as previous studies have not considered more recent drawdown measures, such as the end-of-period drawdown or the new linearly weighted drawdown.

Martin and McCann (1989) to emphasize large losses. To incorporate aspects of regret, the end-of-period drawdown (eopDD) was introduced in Möller (2018).[7]

All these drawdown measures can be subsumed under a unified framework, which we refer to as the *weighted drawdown (wDD) framework* because the main idea is to attach weights to individual drawdowns. Consider an investor who examines the risk of an investment over the period from date 0 to date N and assume that market values $S_0, S_1, \ldots, S_N$ of the investment portfolio are available. Then the wDD is defined as the weighted sum, using weights ω_i, of the drawdowns D_i:

$$\mathrm{wDD} := \sum_{i=1}^{N} \omega_i D_i, \qquad 0 \leq \omega_i \leq 1, \ \sum_{i=1}^{N} \omega_i = 1,$$

where $D_i := \frac{M_i - S_i}{M_i}$ and $M_i := \max\limits_{t=0,\ldots,i} S_t$. The time series of drawdowns D_i is called the drawdown graph. At each point in time, the drawdown graph provides the (percentage) loss incurred from the previous maximum. Different choices of weights ω_i in the wDD framework lead to different drawdown measures and provide valuable information about the properties of these measures. In the following paragraph, we explicitly detail the choices of weights necessary to recover each of the drawdown measures mentioned above.

The ADD is derived by weighting all drawdowns D_i equally, hence setting all weights to $\omega_i = \frac{1}{N}$. To obtain the MDD, the largest element of the drawdown graph receives a weight of one and all of the other elements a weight of zero because only the largest peak to trough loss is considered.[8] The eopDD captures the drawdown at the end of the time period and is defined as the negative return incurred from the time of the global maximum to date N.[9] In terms of ω_i, the weight ω_N equals one and all other weights equal zero. For the ADD^2, where the drawdowns are first squared and then averaged, the weights have to be of the form $\omega_i^* = \frac{1}{N} D_i$. Because these weights do not sum to one, we rescale them by computing $\omega_i = \frac{1}{K}\omega_i^*$ with $K = \sum_{j=1}^{N} \frac{1}{N} D_j$.

[7]The literature contains even more notions of drawdown. Goldberg and Mahmoud (2017) define an ex-ante concept requiring the distribution of MDDs, while Landriault et al. (2015) and Landriault et al. (2017) analyze the frequency and duration of drawdowns of stochastic processes.

[8]All members of the conditional drawdown (CDD) family can be obtained by choosing ω_i as follows: For the CDD at confidence level α, count as n_α all D_i exceeding the α-quantile of the D_is; then $\omega_i = \frac{1}{n_\alpha}$ if D_i exceeds the α-quantile and $\omega_i = 0$ otherwise.

[9]The end of the time period considered by the investor may coincide with the investment horizon or refer to the time at which a regular risk assessment of the investment takes place.

Such rescaling leads to an intuitive interpretation of the weighting scheme because it yields $\omega_i = \frac{D_i}{\sum_{j=1}^{N} D_j}$; that is, each D_i receives a weight proportional to its size against all other D_is.[10]

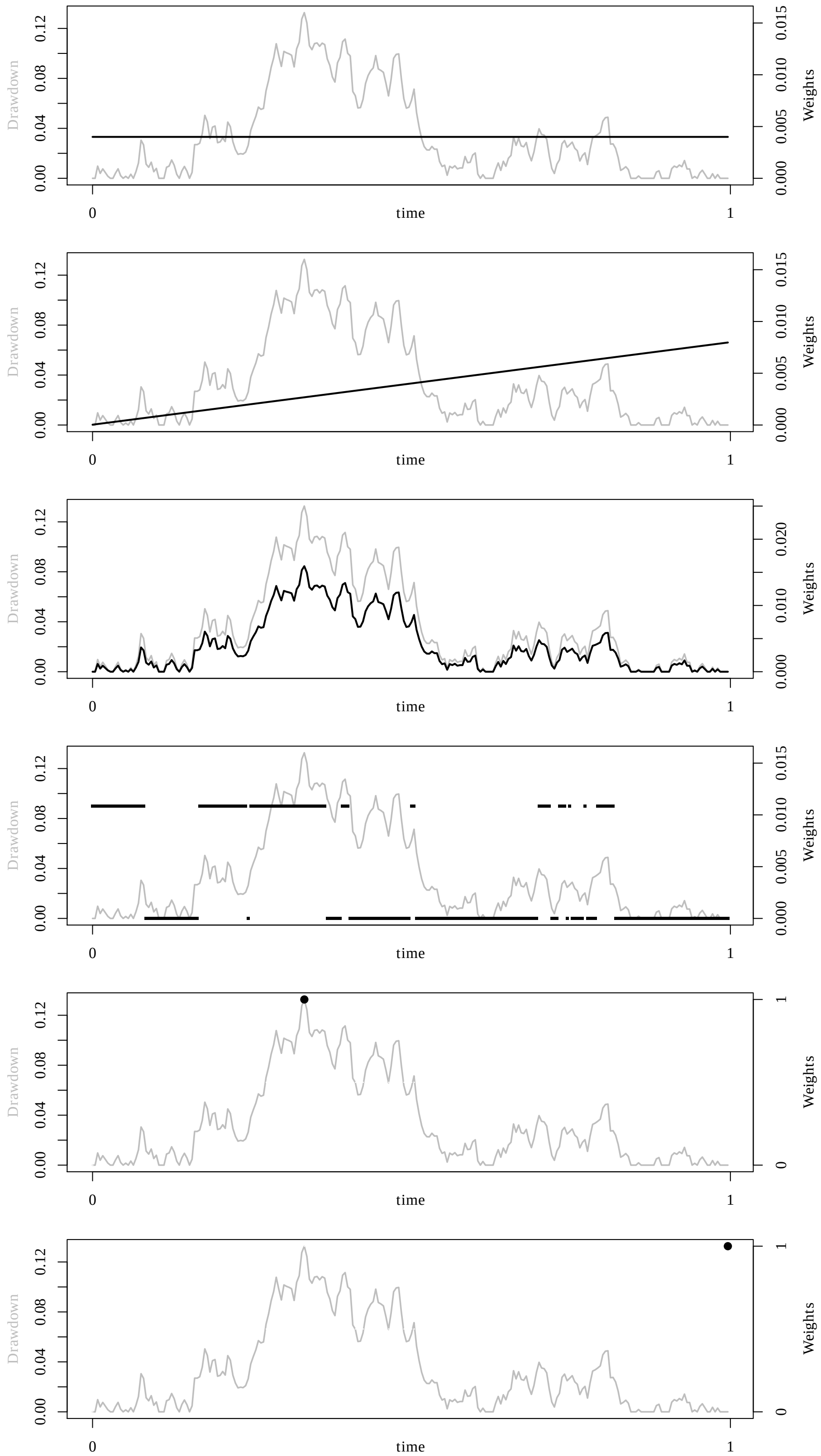
Within the wDD framework, it is easy to design new drawdown measures. The weights can be tailored to an individual's risk preferences, providing an easy way to construct personalized drawdown measures. For example, it may be reasonable that drawdowns toward the end of the time period receive higher weights, as investors may remain calm if drawdowns occur at the beginning, but become increasingly concerned if drawdowns occur toward the end. A set of weights that reflects such preferences is $\omega_i^* = \frac{i}{N}$, where the weights increase linearly from $\frac{1}{N}$ to 1. Since these weights do not sum to one, we rescale them via $\omega_i = \frac{1}{K}\omega_i^*$ where $K = \sum_{j=1}^{N} \frac{j}{N} = \frac{N+1}{2}$. We refer to the resulting drawdown measure as the linearly weighted drawdown (lwDD).

Alternatively, investors may treat drawdowns differently depending on the trend of a strategy. After heavy losses, drawdowns will likely be perceived as being more painful than drawdowns incurred when the strategy is already recovering. To reflect such a pattern, one may set all weights ω_i^* to zero if the strategy's return R_i over the previous month i is positive and to $\frac{1}{N}$ otherwise. For the rescaled version ω_i, the non-zero weights must be chosen as $\frac{1}{N^*}$, with N^* denoting the number of instances where $R_i \leq 0$. The resulting measure—we call it the trend weighted drawdown (twDD)—weights all drawdowns D_i equally but disregards elements of the drawdown graph where the strategy is already recovering.

In fact, many other aspects of risk can be captured within the wDD framework. For example, denote the time from the strategy's last maximum to date i by d_i. Weights chosen as $d_i / \sum_{j=1}^{N} d_j$ attach higher weights to prolonged drawdowns and smaller weights to drawdowns of short duration. Yet another idea is to assign different weights to drawdowns of different intensities. For example, drawdowns below a 5%-threshold may be deemed insignificant and receive weights of zero.

Figure 1 illustrates the weight functions of different drawdown measures for the same (simulated) drawdown graph. It becomes apparent that the weighting schemes differ significantly, highlighting certain similarities and differences between the drawdown measures: While ADD,lwDD and ADD^2 usually attach non-zero weights to most elements of the drawdown graph, MDD and

[10]In the literature, for example in Caporin et al. (2014), the ADD^2 is sometimes defined as the square root of the version defined above. This alternative does not fit into the wDD framework but is a monotonic transformation that does not alter the relative ranking of investments.



Average Drawdown
$$\omega_i = \frac{1}{N}$$

Linearly Weighed Drawdown
$$\omega_i = \frac{2i}{N(N+1)} \propto \frac{i}{N}$$

Average Squared Drawdown
$$\omega_i = \frac{D_i}{\sum_{j=1}^{N} D_j} \propto \frac{1}{N} D_i$$

Trend Weighted Drawdown
$$\omega_i = \begin{cases} \frac{1}{N^*} & \text{if } R_i \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Maximum Drawdown
$$\omega_i = \begin{cases} 1 & \text{if } D_i = \max_{j=1,\ldots,N} D_j \\ 0 & \text{otherwise} \end{cases}$$

End-of-period Drawdown
$$\omega_i = \begin{cases} 1 & \text{if } i = N \\ 0 & \text{otherwise} \end{cases}$$

Figure 1. *Illustration of Different Drawdown Measures within the wDD Framework. Notes: The figure shows the weights ω_i needed to obtain specific drawdown measures (i.e., ADD, lwDD, ADD^2, twDD, MDD, eopDD) within the wDD framework. It uses the same simulated drawdown graph with $N = 250$ for all drawdown measures. This drawdown graph is depicted by the grey lines. The weight functions are shown as black lines or black dots.*

eopDD pick only a single element of the drawdown graph. Depending on the strategy's trend, the number of non-zero weights may vary significantly for the twDD. While the weights of ADD, lwDD and eopDD are predetermined at the beginning of the time period, the weights of ADD^2, twDD and MDD depend on the path of the drawdown graph. While lwDD and eopDD both focus on drawdowns toward the end of the time horizon, MDD and ADD^2 attach the highest weight to the maximum of the drawdown graph. By highlighting different parts of the drawdown graph, each drawdown measure emphasizes different aspects of drawdown.

What does the wDD framework tell us about our main question? Are drawdown measures all the same? Given that many drawdown measures are merely specific versions of the wDD, the drawdown measures could appear to be all the same. However, as Figure 1 shows, the weighting schemes differ markedly from one drawdown measure to the next. This suggests that at least some of the measures differ quite significantly from others. To investigate this issue further, we quantify the degrees of similarity between various drawdown measures in an empirical study.

III Design of Simulation Study

Drawdown measures are applied in many fields[11], most notably in fund management. We analyze drawdown measures within this context by simulating portfolios of fictitious portfolio managers selecting stocks from the MSCI World universe. Unlike a setup under which actual portfolio data (e.g., data from hedge funds or mutual funds for which certain information, including information on the funds' constituents or strategies, may remain confidential or opaque), our setup provides a fully transparent and controlled environment that also allows us to introduce management skill.

Data Sources and Data Processing

For the data period from December 1999 to April 2019, monthly constituents data of the MSCI World index is used to define the investment universe. For each of the constituent stocks, we obtain daily stock prices from Datastream.[12] Any prices denominated in currencies other than

[11]Examples of fields include control theory (Hsieh and Barmish, 2017), insurance (Palmowski and Tumilewicz, 2017), energy markets (Charwand et al., 2017), and option pricing (Dassios and Lim, 2018).

[12]We assume that dividends are reinvested to purchase additional equity. All dividends are on a pre-tax basis.

U.S. dollar (USD) are converted to USD using the spot exchange rate taken from Datastream. To group stocks into broad sectors, we use two-digit Global Industry Classification Standard (GICS) codes.[13] To compute excess returns, U.S. government bond yields for a time to maturity of one year are obtained from Datastream. In total, our investment universe comprises 3,489 stocks from 26 countries and 11 sectors.

Selecting Portfolios

We consider fictitious portfolio managers, who hold portfolios containing 100 stocks picked from the MSCI World index at random. To make the selection process more realistic, we add three design elements. First, we allow for the fact that managers adjust their portfolios over time. Portfolio adjustments occur because stocks leave the index. Moreover, we allow for some additional turnover, leading to a total adjustment of 10% per month.[14] Second, noting that managers not only pick stocks, but also assign a portfolio weight to each stock, we mimic such decisions by assigning each stock a random weight from the set $\{0\%, 0.1\% \ldots, 2\%\}$, such that all weights sum to one. Third, in a realistic setting, managers avoid portfolios that drastically overweight any particular country or sector. Accordingly, we compute the proportions of countries and sectors in the MSCI World index and limit the deviations from each of these proportions to be at most 10 percentage points for each manager's portfolio.

Given these rules for the portfolio selection process, all managers follow the same procedure. On December 31, 1999, they begin by randomly sampling 100 stocks from the index and assigning random weights between 0% and 2%. If the resulting portfolio deviates from the country and sector proportions of the index by more than 10 percentage points, they sample anew until a portfolio satisfies the country and sector bounds. With the current end-of-day stock prices, each manager computes how much of each stock has to be bought to obtain the previously sampled

---

[13] The GICS industry classification codes by MSCI and Standard & Poor's have been found to be superior to other industry classifications, such as the Fama and French Industry Portfolios and the North American Industry Classification System (NAICS) (Hrazdil et al., 2013). They are also widely applied in practice (Scislaw, 2015).

[14] We do not consider transaction costs, which would affect both the costs related to turnover and the subsequent rebalancing of portfolio weights. Assuming proportional transaction costs of 50 bp per transaction, the turnover step would generate annual costs of 1.2%. Transaction costs arising from the rebalancing step depend on the dispersion of asset returns. Accounting for these costs would be arduous and would likely provide no additional insights.

portfolio weights. For each day of the following month, they compute the portfolio values by aggregating the individual stock prices. If a stock price is unavailable, the last obtainable price is used.

At the end of the month, the managers remove all stocks from the portfolio that have exited the index. Additionally, they randomly remove stocks until they arrive at the total monthly turnover of 10%. In the unlikely case that more than 10% of the portfolio's stocks leave the index, there is no additional turnover that month. Next, the managers fill the portfolio back up to 100 stocks by randomly selecting stocks from the new index constituents, excluding those that have been deleted from the portfolio in the previous deletion step. For the new stocks, they also sample new portfolio weights. They draw such sets of new stocks and weights until a new portfolio is found within the country and sector constraints.[15] This procedure is repeatedly applied each month.

Introducing Skill

In the base simulation, all of the managers are treated equally in the sense that their information sets are the same: They all pick from the index constituents purely at random and they all check the country and sector bounds afterwards. We now extend the simulation model and allow managers to have some skill in picking future winners over future losers. When the portfolio is reassembled each month, the universe is split into two halves at the median return of the following year. The upper half outperforms its peers in the following year (by having above median returns), while the lower half underperforms. A skillful manager has the ability to anticipate if a given stock belongs to the upper or lower half, i.e., a skillful manager has some form of foresight. We define different levels of skill by varying a manager's probability to correctly decide if a stock would outperform or underperform. In portfolio management, the above probability is known as the *hit ratio* and shall be denoted by δ.[16] In our simulation study, we vary δ between 50% and 60%, where 50% corresponds to the purely random case and 60%

[15] If in a rare case, this fails in more than 1,000 attempts, the portfolio manager increases the turnover in 1%-steps.

[16] The most prominent skill measure is the α with respect to some factor model. Compared to α, the hit ratio is more immediate because it operates on the level of individual portfolio constituents without further assumptions about the data generating process. It requires information about the investment universe but not about asset pricing factors and factor sensitivities.

aims to model a very skillful manager. Each time a stock is added or deleted from the portfolio during the simulations, we adjust the probabilities according to a manager's hit ratio. Thus, managers with $\delta = 60\%$ skill have a higher chance of including a future winner in their portfolio and of dropping a future loser. On average, such skillful managers generate significantly higher returns at similar levels of standard deviation and skewness (see Table 1). Because of its effect on mean returns, higher manager skill also reduces the expected shortfall and the value-at-risk.

Hit ratio	Mean	SD	Skewness	Min.	Max.	VaR	ES
0.50	0.093	0.213	−0.100	−0.517	0.748	−0.277	−0.422
0.51	0.098	0.213	−0.096	−0.512	0.755	−0.270	−0.416
0.52	0.104	0.213	−0.093	−0.506	0.761	−0.264	−0.410
0.53	0.110	0.213	−0.097	−0.503	0.766	−0.257	−0.406
0.54	0.116	0.213	−0.088	−0.498	0.777	−0.253	−0.401
0.55	0.123	0.213	−0.078	−0.492	0.786	−0.245	−0.394
0.56	0.128	0.214	−0.083	−0.490	0.792	−0.240	−0.392
0.57	0.134	0.214	−0.080	−0.484	0.799	−0.233	−0.385
0.58	0.141	0.214	−0.072	−0.477	0.805	−0.226	−0.378
0.59	0.146	0.214	−0.067	−0.474	0.813	−0.220	−0.374
0.60	0.153	0.214	−0.057	−0.468	0.822	−0.213	−0.368

Table 1. *Summary Statistics of Portfolio Returns for Different Skill Levels. Notes: The summary statistics are computed from annual discrete portfolio returns using a rolling window at a monthly frequency. The numbers represent averages from 1,000 simulated portfolios of fictitious portfolio managers. The value-at-risk and the expected shortfall are computed for the* 95% *confidence level.*

IV Similarity in Portfolio Rankings

As a first aspect of similarity between drawdown measures, we investigate whether these measures lead to the same ranking of portfolios. As a measure of similarity, we use rank correlations between portfolio rankings resulting from different drawdown measures.[17] We conduct pairwise comparisons for all measures from Section II, and include the standard deviation and the ex-

[17]To quantify the degree of similarity between the measures, rank correlations have been widely applied (see, for example, Eling (2008), Haas Ornelas et al. (2012), or Auer and Schuhmacher (2013)).

pected shortfall for comparison.[18] The objects to be ranked are the portfolios of 1,000 fictitious portfolio managers, as described in Section III. Rank correlations are obtained using Kendall's τ.[19] For each of the 210 (overlapping) one-year periods[20] in our data period, we compute the rank correlation between two risk measures and report the time-averaged rank correlations in the lower triangle of Table 2. The upper triangle contains 99% confidence intervals for the corresponding averages computed by using Newey-West standard errors with 12 lags. Panel A of Table 2 contains the results for the purely random case ($\delta = 0.5$) and Panel B contains the results for highly skillful managers ($\delta = 0.6$). For hit ratios in between, the rank correlations are well-behaved and tend to descend from the values in Panel A to the values in Panel B.

The results in Table 2 have a straightforward interpretation. When someone compares the 1,000 unskillful portfolio managers using the ADD and someone compares them using the lwDD during a one year period, both rankings would (on average) be correlated by 0.840 (see Table 2, Panel A, top left entry). As this correlation is fairly close to one, ADD and lwDD appear to be fairly similar to each other.

In terms of general structures, Panel A of Table 2 shows high pairwise correlations of 0.85 ± 0.03 between ADD, ADD^2 and lwDD. Thus, these three measures produce very similar rankings. These results are highly plausible in light of the wDD framework, as illustrated in Figure 1, because all three drawdown measures assign weights to all elements of the drawdown graph. Correlations of these measures with twDD, which assigns weights to varying parts of the drawdown graph, drop slightly, whereas correlations with MDD and eopDD, with only a single

[18]The expected shortfall is computed for the 95% confidence level. In our setup, employing the expected shortfall or the value-at-risk leads to virtually the same results.

[19]Kendall's τ and Spearman's ρ are the most common choices for rank correlation measures. While Spearman's ρ lacks a straightforward interpretation, Kendall's τ can easily be interpreted as the probability of two pairs of observations being concordant minus the probability of being discordant (Noether, 1981). In our setting, it is advisable to use version b) of Kendall's τ, which corrects for tied ranks, because the eopDD is frequently zero, which leads to tied ranks.

[20]We employ data from December 31, 1999 to April 30, 2019. To implement different hit ratios, one year of future data is required after a portfolio is set up. Thus, we set up the last portfolio on April 30, 2018. If we wanted to update the portfolio one month later, data beyond our data period were necessary. Since all drawdown measures are evaluated over one-year intervals, the end of the last evaluation period is May 31, 2018. This leads to a total of 210 one-year periods.

Panel A: Managers without skill (hit ratio 0.5)

	ADD	lwDD	ADD^2	twDD	MDD	eopDD	ES	SD
ADD	1	$[0.81, 0.87]$	$[0.85, 0.90]$	$[0.68, 0.79]$	$[0.54, 0.63]$	$[0.21, 0.43]$	$[0.27, 0.33]$	$[0.27, 0.33]$
lwDD	0.840	1	$[0.77, 0.87]$	$[0.62, 0.76]$	$[0.51, 0.63]$	$[0.27, 0.50]$	$[0.25, 0.31]$	$[0.24, 0.31]$
ADD^2	0.874	0.821	1	$[0.76, 0.83]$	$[0.63, 0.70]$	$[0.23, 0.47]$	$[0.29, 0.35]$	$[0.28, 0.35]$
twDD	0.736	0.690	0.797	1	$[0.58, 0.65]$	$[0.21, 0.45]$	$[0.27, 0.34]$	$[0.27, 0.35]$
MDD	0.586	0.568	0.668	0.617	1	$[0.19, 0.43]$	$[0.32, 0.41]$	$[0.32, 0.41]$
eopDD	0.323	0.387	0.351	0.329	0.311	1	$[0.11, 0.22]$	$[0.11, 0.21]$
ES	0.299	0.281	0.321	0.308	0.367	0.165	1	$[0.60, 0.69]$
SD	0.298	0.275	0.314	0.309	0.366	0.157	0.644	1

Panel B: Managers with significant skill (hit ratio 0.6)

	ADD	lwDD	ADD^2	twDD	MDD	eopDD	ES	SD
ADD	1	$[0.80, 0.86]$	$[0.85, 0.89]$	$[0.66, 0.78]$	$[0.53, 0.62]$	$[0.17, 0.37]$	$[0.27, 0.33]$	$[0.26, 0.33]$
lwDD	0.828	1	$[0.75, 0.85]$	$[0.60, 0.73]$	$[0.48, 0.61]$	$[0.23, 0.44]$	$[0.25, 0.31]$	$[0.24, 0.31]$
ADD^2	0.867	0.801	1	$[0.75, 0.83]$	$[0.63, 0.70]$	$[0.19, 0.41]$	$[0.29, 0.36]$	$[0.28, 0.35]$
twDD	0.720	0.665	0.787	1	$[0.57, 0.64]$	$[0.17, 0.38]$	$[0.27, 0.35]$	$[0.26, 0.35]$
MDD	0.572	0.547	0.660	0.607	1	$[0.15, 0.36]$	$[0.32, 0.41]$	$[0.31, 0.40]$
eopDD	0.272	0.332	0.296	0.275	0.258	1	$[0.10, 0.20]$	$[0.09, 0.19]$
ES	0.300	0.280	0.322	0.307	0.361	0.147	1	$[0.60, 0.69]$
SD	0.296	0.272	0.312	0.305	0.355	0.140	0.642	1

Table 2. *Rank Correlations Between ADD, lwDD, ADD^2, twDD, MDD, eopDD, Expected Shortfall (ES) and Standard Deviation (SD). Notes: The lower triangle contains the average rank correlation of the portfolio rankings (average over 210 one-year periods); the upper triangle contains the corresponding 99% confidence intervals. Panel A reports the results for a hit ratio of 0.5 and Panel B reports the results for a hit ratio of 0.6.*

element receiving non-zero weights, drop sharply. Furthermore, it is not surprising that MDD has ADD^2 as its closest relative because MDD and ADD^2 assign the highest weight ω_i to the same element of the drawdown graph. The eopDD exhibits the lowest rank correlations to its peers, indicating that its rankings are significantly different from those of the other drawdown measures. Its closest relative is lwDD, which also allocates the highest weight to the last element of the drawdown graph. Moreover, for all pairs of risk measures, the rank correlation is strictly positive and below 0.9 at the 99% confidence level. In comparison, expected shortfall and standard deviation are rather weakly correlated with the drawdown measures. With a range of rank correlations between 0.258 and 0.874, we can conclude that some of the six drawdown measures under consideration are closely related while others are very different.

A comparison between Panels A and B shows that rank correlations between drawdown measures are very robust with respect to the hit ratio. Changes are largest for the eopDD where correlations decrease monotonically as skill increases. This finding reflects the eopDD's particular sensitivity to changes in the first return moment (Möller, 2018) that occur when skill changes.

V Similarity in Skill Detection

A second aspect of similarity between drawdown measures is whether they are equally adept at differentiating between skillful and unskillful managers. Managerial skill and drawdown should be interconnected: Managers with high hit ratios select more future winners than future losers. Consequently, their portfolios should experience lower drawdowns because drawdowns are a direct consequence of losses. If high hit ratios lead to lower drawdowns, relatively low drawdowns may be used as an indicator of investor skill. Therefore, we want to explore whether all drawdown measures are equally suited for this purpose.

To test similarity in skill detection, we consider a setting in which 1,000 skillful managers with a hit ratio of 60% and 1,000 unskillful managers with a hit ratio of 50% manage their portfolios for a given year. At the end of the year, we observe the 2,000 portfolio paths and try to distinguish between the skillful and unskillful managers. As we expect the skillful managers to have lower drawdowns, our best guess is that all portfolios below the median drawdown belong to a skillful manager and all portfolios above the median drawdown belong to an unskillful manager. If the relationship between drawdown and skill were strong, close to 100% of all 2,000 managers would be classified correctly; not a single correct classification would indicate a strong opposite effect. If drawdowns and skill were independent, the classification would be approximately 50% accurate by chance. We use this percentage as a discrimination measure to compare how well different drawdown measures detect skill.

Figure 2 shows how accurately the drawdown measures from Section II discriminate between managers with more and less skill. The standard deviation is also considered for comparison. To arrive at the boxplots in Figure 2, we use a one year rolling window with monthly steps,

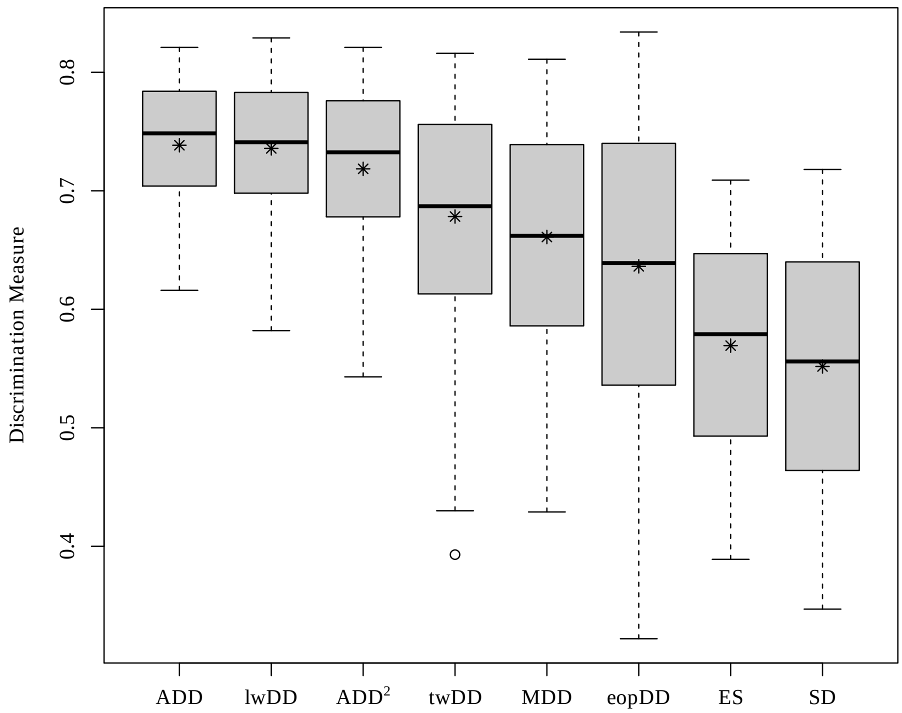


Figure 2. *Discrimination Between Skillful and Unskillful Portfolio Managers Using Drawdown Measures of Risk. Notes: Each boxplot illustrates how the corresponding risk measure discriminates between skillful and unskillful managers. Each month for approximately 20 years of data, we observe the performance of 1,000 skillful managers and 1,000 unskillful managers over the preceding year and classify the managers based on the drawdown measures. The proportion of correctly classified managers is reported on the y-axis. The asterisk additionally depicts the average discrimination measure (across our 210 observations).*

resulting in 210 yearly periods. For each yearly period, we determine how many managers are classified correctly by the different drawdown measures. Thus, the boxplots are indicators of how accurately each drawdown measure detects skill over time and across market phases. The asterisk reports the average discrimination measure. For example, during our historical one-year-periods, ADD classified at best 82% and at worst 62% of managers correctly with an average of 74%.

We find that ADD performs best and exhibits significant skill detection abilities. Moreover, the dispersion over time is smallest in comparison. lwDD is almost as successful as ADD, followed by ADD^2. In contrast, twDD, MDD and eopDD are much less accurate at detecting skill—both on average and with respect to dispersion over time. During some periods, their classifications are helpful; during other periods, the classifications are worse than random. To compare different drawdown measures, the relative performance is key because the absolute levels of our discrimination measure depend on the difference in hit ratios, which is 10% for the results in Figure 2. However, variation in this difference shows that the relative performance of the various drawdown measures remains unchanged. Although the standard deviation exhibits the poorest performance in detecting skill, it leads to discrimination measures slightly above 0.5 on average, indicating some discriminatory power. The reason for this could be a negative correlation between stock returns and volatility in the cross section due to the leverage effect (Black, 1976) or the volatility feedback effect (Campbell and Hentschel, 1992).

In summary, drawdown measures are useful in discriminating between skillful and unskillful managers. Their tendency to penalize losses gives them a more holistic view than risk measures that only penalize variability. We also find significant differences between the drawdown measures. Specifically, measures that incorporate more information about the drawdown graph (i.e., ADD, lwDD and ADD^2) outperform those measures that focus on fewer elements of the graph (i.e., twDD, MDD and eopDD). All are markedly better than the expected shortfall and the standard deviation, which only have little power to discriminate between skillful and unskillful managers.

VI Drawdown-based Performance Ratios

Apart from their immediate application as a risk measure, drawdown measures are employed in performance measurement. A common technique for constructing performance measures is to divide excess returns (over the risk-free rate) by some measure of risk (Caporin et al., 2014).[21] Some of the resulting drawdown-based performance ratios are already known from the literature and have received names of their own; for example, *Calmar ratio* when MDD is in the denominator, *Pain ratio* when ADD is in the denominator, and *Ulcer ratio* or *Martin ratio* when ADD^2 is in the denominator (Cogneau and Hübner, 2009; Schuhmacher and Eling, 2011). For clarity, we do not use the names of these measures in the following, but we denote the drawdown-based performance ratios like the corresponding drawdown measures with an

[21]This plug-in technique used to construct performance ratios may be theoretically sound for some risk measures in some circumstances, like the Sharpe ratio under normally distributed returns. However, there is no theoretical justification to apply this technique to drawdown measures. Similar *ad hoc* measures have also been criticized by Leland (1999).

additional subscript r.

For the final step of our empirical analysis, we investigate similarities between different drawdown-based performance ratios. Again, we examine similarities in terms of portfolio rankings and the ability to detect portfolio managers' skill. To investigate similarities in portfolio rankings, we compute the rank correlation for each pair of drawdown ratios. Table 3 sets out the results of our analysis. We find that all correlations fall between 0.443 and 0.937, which is substantially higher than the correlations obtained when portfolio rankings are based on drawdown risk measures (compare Table 2). While correlations with the $\text{eopDD}_\mathbf{r}$ are still comparatively low at approximately 0.47, correlations with the Sharpe ratio have increased substantially to approximately 0.77. The general increase in rank correlations is likely due to the common numerator of all performance ratios, the excess return. Nevertheless, all patterns present in Table 2 remain intact and the intuitions of the wDD framework remain valid. For example, $\text{ADD}_\mathbf{r}$ is still closest related to $\text{ADD}^2_\mathbf{r}$ followed by $\text{lwDD}_\mathbf{r}$, $\text{twDD}_\mathbf{r}$ and $\text{MDD}_\mathbf{r}$ and a wide gap to $\text{eopDD}_\mathbf{r}$. In conclusion, we find that drawdown-based performance ratios are more similar than the drawdown measures themselves, but still display important differences that are in line with the wDD framework.

Finally, we investigate the extent to which drawdown performance ratios are similar at detecting skill. Figure 3 shows the proportions of correctly classified portfolio managers. This figure reveals several interesting findings. First, compared to the corresponding results for the drawdown measures in Figure 2, drawdown-based performance ratios are better at detecting skill on average. This finding is expected, since ratios use more information due to the excess return in the numerator. Second, all drawdown-based performance ratios perform similarly well, except for $\text{eopDD}_\mathbf{r}$, which performs considerably worse. This is likely due to the following property of $\text{eopDD}_\mathbf{r}$. In cases in which the time series of portfolio values reaches its maximum at the end, eopDD is zero. Mathematically, dividing by zero is infeasible in the ratio; economically, this case constitutes the optimal 'no risk' outcome. We resolve this issue in the economically sensible way by treating all managers with zero eopDD as equally and infinitely good; however, this leads to a significant number of ties. Third, the Sharpe ratio appears to detect skill more accurately than the drawdown-based performance ratios. This is quite surprising, as the drawdown measures themselves already have some ability to detect skill, much more so than the

Panel A: Managers without skill (hit ratio 0.5)

	ADD_r	$lwDD_r$	ADD_r^2	$twDD_r$	MDD_r	$eopDD_r$	ES_r	SD_r
ADD_r	1	$[0.89, 0.93]$	$[0.92, 0.95]$	$[0.85, 0.91]$	$[0.74, 0.87]$	$[0.34, 0.57]$	$[0.67, 0.83]$	$[0.70, 0.85]$
$lwDD_r$	0.909	1	$[0.88, 0.93]$	$[0.82, 0.88]$	$[0.72, 0.85]$	$[0.37, 0.59]$	$[0.65, 0.81]$	$[0.67, 0.82]$
ADD_r^2	0.937	0.903	1	$[0.88, 0.93]$	$[0.79, 0.90]$	$[0.36, 0.59]$	$[0.69, 0.84]$	$[0.71, 0.85]$
$twDD_r$	0.882	0.850	0.909	1	$[0.77, 0.88]$	$[0.35, 0.59]$	$[0.67, 0.82]$	$[0.69, 0.83]$
MDD_r	0.804	0.785	0.843	0.821	1	$[0.36, 0.59]$	$[0.73, 0.85]$	$[0.75, 0.86]$
$eopDD_r$	0.452	0.482	0.473	0.472	0.473	1	$[0.34, 0.57]$	$[0.34, 0.58]$
ES_r	0.753	0.732	0.763	0.748	0.792	0.455	1	$[0.89, 0.94]$
SD_r	0.772	0.748	0.778	0.761	0.803	0.458	0.912	1

Panel B: Managers with significant skill (hit ratio 0.6)

	ADD_r	$lwDD_r$	ADD_r^2	$twDD_r$	MDD_r	$eopDD_r$	ES_r	SD_r
ADD_r	1	$[0.88, 0.93]$	$[0.92, 0.96]$	$[0.84, 0.91]$	$[0.74, 0.87]$	$[0.31, 0.58]$	$[0.69, 0.84]$	$[0.72, 0.85]$
$lwDD_r$	0.908	1	$[0.87, 0.93]$	$[0.80, 0.89]$	$[0.72, 0.85]$	$[0.34, 0.60]$	$[0.67, 0.82]$	$[0.69, 0.83]$
ADD_r^2	0.937	0.897	1	$[0.87, 0.93]$	$[0.80, 0.90]$	$[0.32, 0.59]$	$[0.70, 0.84]$	$[0.72, 0.85]$
$twDD_r$	0.876	0.842	0.904	1	$[0.76, 0.88]$	$[0.32, 0.59]$	$[0.67, 0.82]$	$[0.69, 0.83]$
MDD_r	0.808	0.787	0.847	0.82	1	$[0.32, 0.59]$	$[0.71, 0.85]$	$[0.72, 0.85]$
$eopDD_r$	0.443	0.469	0.459	0.455	0.455	1	$[0.31, 0.58]$	$[0.31, 0.58]$
ES_r	0.763	0.742	0.767	0.744	0.778	0.444	1	$[0.87, 0.93]$
SD_r	0.785	0.761	0.783	0.758	0.788	0.448	0.904	1

Table 3. *Rank Correlations between ADD_r, $lwDD_r$, ADD_r^2, $twDD_r$, MDD_r, $eopDD_r$, ES_r, and the Sharpe Ratio. Notes: The lower triangle contains the average rank correlation of the portfolio rankings (average over 210 one-year periods); the upper triangle contains the corresponding* 99% *confidence intervals. Panel A reports the results for a hit ratio of 0.5 and Panel B reports the results for a hit ratio of 0.6.*

standard deviation. Fourth, somewhat surprisingly, ADD_r, $lwDD_r$, $twDD_r$, ADD_r^2, and $eopDD_r$ discriminate markedly worse than chance in some subperiods.

The third and fourth findings require further explanation. When we examine the cases when drawdown-based performance ratios classify particularly badly, we see that significant portfolio losses strongly affect both the ratio's numerator and denominator, such that when the return in the numerator becomes more negative, the drawdown measure strongly increases. For example, a relatively skillful portfolio manager with a loss of 42% and ADD of 0.14 might end up with a worse ratio (-3) than an unskillful manager with a (higher!) loss of 50% and a (worse!) ADD of 0.20 but a ratio of -2.5. This is just one example of a more general effect. In the appendix,

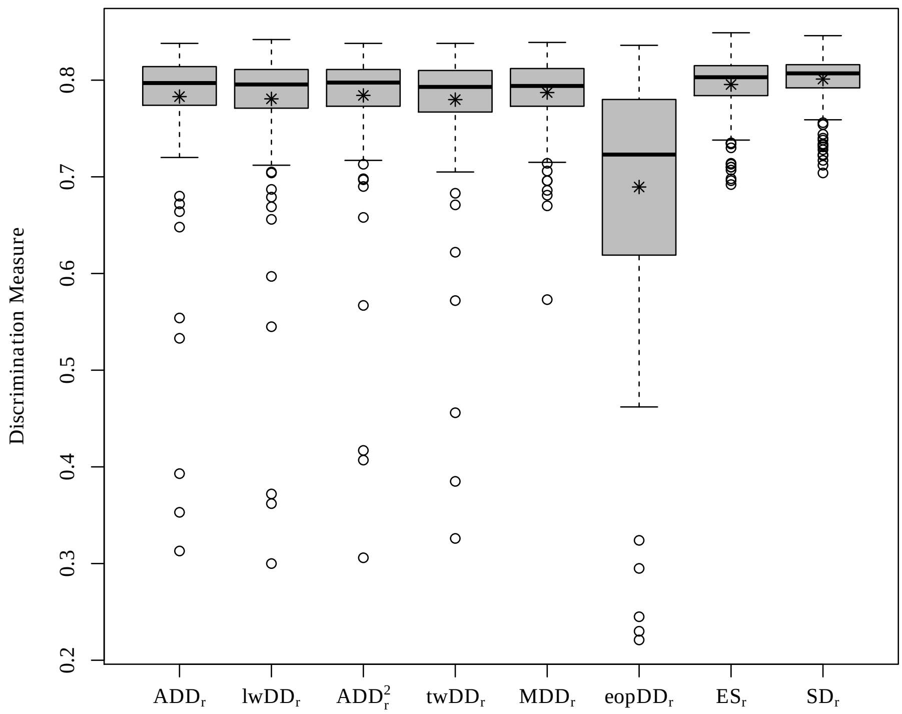


Figure 3. *Discrimination Between Skillful and Unskillful Portfolio Managers Using Drawdown-based Performance Ratios. Notes: Each boxplot illustrates how the corresponding performance ratio discriminates between skillful and unskillful managers. Each month for approximately 20 years of data, we observe the performance of 1,000 skillful managers and 1,000 unskillful managers over the preceding year and classify the managers based on the drawdown measures. The proportion of correctly classified managers is reported on the y-axis. The asterisk additionally depicts the average discrimination measure (across our 210 observations).*

we characterize all of the cases in which unskillful managers obtain higher ratios than skillful managers. The Sharpe ratio is not immune to this effect either.[22] However, drawdown ratios are particularly susceptible to this effect, as the numerator and denominator of the ratio are closely interrelated. By definition, drawdown measures capture losses from a running maximum, which typically occur when prices are falling and returns are negative. As positive returns increase and drawdowns tend to be small, the ratio becomes large. As negative returns fall further and drawdowns typically spike, the ratio may remain unaltered because both effects offset each other. One strength of drawdown measures is that they are particularly alert to losses; however, this strength may become a drawback when plugging them into ratios with the excess returns in the numerator. Therefore, we conclude that a naïve application of drawdown measures in performance ratios may not be particularly useful overall.

VII Conclusion

Drawdown measures provide a number of practical and theoretical benefits: They are intuitive path-dependent risk measures, which focus on downside risk and capture psychologically important aspects, such as regret. Consequently, it is no surprise that different variants of these measures have been developed in the past and widely applied in practice.

We establish that most of the existing measures can be summarized under the wDD framework. Moreover, new measures capturing investor-specific preferences can easily be developed within the framework. This theoretical insight may provide guidance for choosing the most appropriate drawdown measure for one's own purposes. However, an immediate question arises: If all these measures fit into the same framework, are they all fundamentally the same and do they all lead to the same conclusions? To answer this question, we investigate the similarity of drawdown measures empirically in two applications: the ranking of portfolios and the ability to detect the skill of portfolio managers. Our results show that drawdown measures are certainly not all the same. Moreover, observed similarities and differences between the drawdown measures are well

[22] This "perverse" effect is acknowledged in the literature at least since Jobson and Korkie (1981). The ensuing debate on whether the Sharpe ratio should be used when returns are negative has led to numerous contributions in its favor, for example, Sharpe (1998), Akeda (2003), McLeod and van Vuuren (2004), and against, for example, Ferruz Agudo and Sarto Marzal (2004), Israelsen (2005) and Scholz (2007).

in line with the intuition from the wDD framework.

Since risk measures are also commonly used in performance measurement, we finally explore the similarity of drawdown-based performance ratios. We find that such ratios produce more similar portfolio rankings than the drawdown measures themselves; however, there are still some differences between ratios. In detecting skill, drawdown-based performance ratios perform well on average but poorly in periods of negative returns. While similar ramifications affect other performance ratios, the effect on drawdown-based performance ratios is even worse, which questions the naïve application of drawdown measures for this application.

Appendix

In this appendix, we characterize when a more skillful manager (with higher returns and lower risk) obtains a worse performance ratio than a less skillful manager (with lower returns and higher risk). Let ret_1, risk_1 and ratio_1 denote the return, risk and performance ratio of the skillful manager, respectively, and denote the corresponding quantities of the less skillful manager with index 2. It should be noted that for all risk measures under consideration $\mathrm{risk}_i \geq 0$. We express $\mathrm{ret}_2 = \mathrm{ret}_1 \cdot \alpha$ and $\mathrm{risk}_2 = \mathrm{risk}_1 \cdot \beta$. Since we want to characterize when the skillful manager obtains the worse ratio despite having higher returns and lower risk, i.e., when $\mathrm{ratio}_1 < \mathrm{ratio}_2$, $\mathrm{ret}_1 > \mathrm{ret}_2$ and $\mathrm{risk}_1 < \mathrm{risk}_2$, we always have $\beta > 1$. We distinguish between three distinct cases:

(i) If $\mathrm{ret}_1 < 0$, both returns are negative and $\alpha > 1$. Hence, the skillful manager has the lower ratio if and only if

$$\mathrm{ratio}_1 < \mathrm{ratio}_2 \iff \frac{\mathrm{ret}_1}{\mathrm{risk}_1} < \frac{\mathrm{ret}_2}{\mathrm{risk}_2} \iff \frac{\mathrm{ret}_1}{\mathrm{risk}_1} < \frac{\mathrm{ret}_1 \cdot \alpha}{\mathrm{risk}_1 \cdot \beta} \iff 1 > \frac{\alpha}{\beta}.$$

Note that the sign changes in the last step because ret_1 is negative. Hence, the ratio misrepresents the investors' skill if $\beta > \alpha$; that is, when the less skillful manager has a higher relative difference in risk than return compared to a more skillful manager.

(ii) If $\mathrm{ret}_2 > 0$, both returns are positive and $\alpha < 1$. Analogously,

$$\mathrm{ratio}_1 < \mathrm{ratio}_2 \iff \frac{\mathrm{ret}_1}{\mathrm{risk}_1} < \frac{\mathrm{ret}_2}{\mathrm{risk}_2} \iff \frac{\mathrm{ret}_1}{\mathrm{risk}_1} < \frac{\mathrm{ret}_1 \cdot \alpha}{\mathrm{risk}_1 \cdot \beta} \iff 1 < \frac{\alpha}{\beta}$$

because ret_1 is positive. Since $\alpha < 1$ and $\beta > 1$ the condition $\beta < \alpha$ is never attainable.

(iii) If $\mathrm{ret}_1 > 0 > \mathrm{ret}_2$, then $\alpha < 0$. As in the previous case, $\beta < \alpha$ is never satisfied because $\alpha < 0$ and $\beta > 1$.

In summary, a more skillful manager will obtain a worse ratio despite having superior risk and return if and only if both returns are negative and $\beta > \alpha$; that is, when the managers' risks differ more than the returns.

References

Akeda, Y. (2003). Another Interpretation of Negative Sharpe Ratio. *Journal of Performance Measurement 7*(3), 19–23.

Auer, B. R. (2015). Does the Choice of Performance Measure Influence the Evaluation of Commodity Investments? *International Review of Financial Analysis 38*, 142–150.

Auer, B. R. and F. Schuhmacher (2013). Robust Evidence on the Similarity of Sharpe Ratio and Drawdown-Based Hedge Fund Performance Rankings. *Journal of International Financial Markets, Institutions and Money 24*(1), 153–165.

Berk, J. B. and J. H. van Binsbergen (2015). Measuring Skill in the Mutual Fund Industry. *Journal of Financial Economics 118*(1), 1–20.

Black, F. (1976). Studies of Stock Price Volatility Changes. *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association*, 177–181.

Burke, G. (1994). A Sharper Sharpe Ratio. *Futures 23*(3), 56.

Campbell, J. Y. and L. Hentschel (1992). No News is Good News. An Asymmetric Model of Changing Volatility in Stock Returns. *Journal of Financial Economics 31*(3), 281–318.

Caporin, M., G. M. Jannin, F. Lisi, and B. B. Maillet (2014). A Survey on the Four Families of Performance Measures. *Journal of Economic Surveys 28*(5), 917–942.

Caporin, M. and F. Lisi (2011). Comparing and Selecting Performance Measures Using Rank Correlations. *Economics: The Open-Access, Open-Assessment E-Journal 5*.

Charwand, M., M. Gitizadeh, and P. Siano (2017). A New Active Portfolio Risk Management for an Electricity Retailer Based on a Drawdown Risk Preference. *Energy 118*, 387–398.

Chekhlov, A., S. Uryasev, and M. Zabarankin (2005). Drawdown Measure in Portfolio Optimization. *International Journal of Theoretical and Applied Finance 8*(1), 13–58.

Cogneau, P. and G. Hübner (2009). The 101 Ways to Measure Portfolio Performance. *Working Paper, University of Liège*.

Dassios, A. and J. W. Lim (2018). An Efficient Algorithm for Simulating the Drawdown Stopping Time and the Running Maximum of a Brownian Motion. *Methodology and Computing in Applied Probability 20*(1), 189–204.

Eling, M. (2008). Does the Measure Matter in the Mutual Fund Industry? *Financial Analysts Journal 64*(3), 1–13.

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. and K. R. French (2010). Luck versus Skill in the Cross-Section of Mutual Fund Returns. *Journal of Finance 65*(5), 1915–1947.

Ferruz Agudo, L. and J. L. Sarto Marzal (2004). An Analysis of Spanish Investment Fund Performance: Some Considerations Concerning Sharpe's Ratio. *Omega - The International Journal of Management Science 32*(4), 273–284.

Frydman, C. and C. Camerer (2016). Neural Evidence of Regret and its Implications for Investor Behavior. *Review of Financial Studies 29*(11), 3108–3139.

Garcia, C. B. and F. J. Gould (1987). A Note on the Measurement of Risk in a Portfolio. *Financial Analysts Journal 43*(2), 61–69.

Goldberg, L. R. and O. Mahmoud (2017). Drawdown: From Practice to Theory and Back Again. *Mathematics and Financial Economics 11*(3), 275–297.

Grinblatt, M. and S. Titman (1989). Mutual Fund Performance – An Analysis of Quarterly Portofolio Holdings. *Journal of Business 62*(3), 393–416.

Haas Ornelas, J. R., A. F. Silva Júnior, and J. L. Barros Fernandes (2012). Yes, the Choice of Performance Measure does Matter for Ranking of US Mutual Funds. *International Journal of Finance & Economics 17*, 61–72.

Hrazdil, K., K. Trottier, and R. Zhang (2013). A Comparison of Industry Classification Schemes: A Large Sample Study. *Economics Letters 118*(1), 77–80.

Hsieh, C. H. and B. R. Barmish (2017). On Drawdown-Modulated Feedback Control in Stock Trading. *IFAC-PapersOnLine 50*(1), 952–958.

Israelsen, C. (2005). A Refinement to the Sharpe Ratio and Information Ratio. *Journal of Asset Management 5*, 423–427.

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

Jobson, J. D. and B. M. Korkie (1981). Performance Hypothesis Testing with the Sharpe and Treynor Measures. *Journal of Finance 36*(4), 889–908.

Kestner, L. N. (1996). Getting a Handle on True Performance. *Futures 25*(1), 44–46.

Landriault, D., B. Li, and H. Zhang (2015). On the Frequency of Drawdowns for Brownian Motion Processes. *Journal of Applied Probability 52*(1), 191–208.

Landriault, D., B. Li, and H. Zhang (2017). On Magnitude, Asymptotics and Duration of Drawdowns for Lévy Models. *Bernoulli 23*(1), 432–458.

Leland, H. (1999). Beyond Mean-Variance: Performance Measurement in a Nonsymmetrical World. *Financial Analysts Journal 55*, 27–36.

Martin, P. G. and B. B. McCann (1989). *The Investor's Guide to Fidelity Funds.* London: Wiley.

McLeod, W. and G. van Vuuren (2004). Interpreting the Sharpe Ratio when Excess Returns are Negative. *Investment Analysts Journal 33*(59), 15–20.

Möller, P. M. (2018). Drawdown Measures and Return Moments. *International Journal of Theoretical and Applied Finance 21*(7), 1850042.

Noether, G. E. (1981). Why Kendall Tau? *Teaching Statistics 3*(2), 41–43.

Odo, M. (2006). Pain Index and Pain Ratio. *Zephyr Associates White Paper*.

Palmowski, Z. and J. Tumilewicz (2017). Pricing Insurance Drawdown-Type Contracts with Underlying Lévy Assets. *Insurance: Mathematics and Economics 79*, 1–14.

Schmielewski, F. and C. Schwehm (2014). What is Risk? *Journal of Indexes November/December 2014*, 70–79.

Scholz, H. (2007). Refinements to the Sharpe Ratio: Comparing Alternatives for Bear Markets. *Journal of Asset Management 7*(5), 347–357.

Schuhmacher, F. and M. Eling (2011). Sufficient Conditions for Expected Utility to Imply Drawdown-based Performance Rankings. *Journal of Banking & Finance 35*(9), 2311–2318.

Scislaw, K. E. (2015). The Value Premium Within and Across GICS Industry Sectors in a Pre-financial Collapse Sample. *Cogent Economics and Finance 3*(1), 1–18.

Sharpe, W. F. (1998). Morningstar's Risk-Adjusted Ratings. *Financial Analysts Journal 51*(4), 21–33.

Young, T. W. (1991). Calmar Ratio: A Smoother Tool. *Futures 20*(1), 40.

cfr working paper series

centre for financial research cologne

cfr working papers are available for download from www.cfr-cologne.de.

2019

No.	Author(s)	Title
19-04	O. Korn, P. M. Möller, Ch. Schwehm	Drawdown Measures: Are They All the Same?
19-03	G. Cici, M. Hendriock, S. Jaspersen, A. Kempf	#MeToo Meets the Mutual Fund Industry: Productivity Effects of Sexual Harassment
19-02	T. Johann, S. Scharnowski, E. Theissen, C. Westheide, L. Zimmermann	Liquidity in the German Stock Market
19-01	A. Betzer, P. Limbach, P. Raghavendra Rau, Henrik Schürmann	Till death (or divorce) do us part: Early-life family disruption and fund manager behavior

2018

No.	Author(s)	Title
18-06	M. Goergen, P. Limbach, M. Scholz-Daneshgari	On the Choice of CEO Duality: Evidence from a Mandatory Disclosure Rule
18-05	F. Brochet, P. Limbach, D. Bazhutov, A. Betzer, M. Doumet	Where do Investor Relations Matter the Most?
18-04	G. Cici, A. Kempf, C. Peitzmeier	Knowledge Spillovers in the Mutual Fund Industry through Labor Mobility
18-03	T. Rischen, E. Theissen	Underpricing in the Euro Area Corporate Bond Market: New Evidence from Post-Crisis Regulation and Quantitative Easing
18-02	S. Lesmeister, P. Limbach, M. Goergen	Trust and Shareholder Voting
18-01	G. Cici, M. Hendriock, A. Kempf	The Impact of Labor Mobility Restrictions on Managerial Actions: Evidence from the Mutual Fund Industry

2017

No.	Author(s)	Title
17-04	G. Cici, P.B. Shane, Y. S. Yang	Do Connections with Buy-Side Analysts Inform Sell-Side Analyst Research?
17-03	G. Cici, S. Gibson, R. Moussawi	Explaining and Benchmarking Corporate Bond Returns

17-02	S. Jaspersen, P. Limbach	Knowing Me, Knowing You? Similarity to the CEO and Fund Managers' Investment Decisions
17-01	J. Grammig, E.-M. Küchlin	A two-step indirect inference approach to estimate the long-run risk asset pricing model

2016

No.	Author(s)	Title
16-12	A. Betzer, H.S. Lee, P. Limbach, J.M. Salas	Are Generalists Beneficial to Corporate Shareholders? Evidence from Exogenous Executive Turnovers
16-11	F. Brochet, P. Limbach, M. Schmid, M. Scholz-Daneshgari	CEO tenure and firm value
16-10	V. Agarwal, R. Vashishtha, M. Venkatachalam	Mutual fund transparency and corporate myopia
16-09	M.-A. Göricke	Do Generalists Profit from the Fund Families' Specialists? Evidence from Mutual Fund Families Offering Sector Funds
16-08	S. Kanne, O. Korn, M.Uhrig-Homburg	Stock Illiquidity, Option Prices and Option Returns
16-07	S. Jaspersen	Betting on Market Power: Product Market Competition and Mutual Funds' Information Production
16-06	O. Korn, M.-O. Rieger	Hedging With Regret
16-05	E. Theissen, C. Westheide	Call of Duty: Designated Market Maker Participation in Call Auctions
16-04	P. Gomber, S. Sagade, E. Theissen, M.C. Weber, C. Westheide	Spoilt for Choice: Order Routing Decisions in Fragmented Equity Markets
16-03	T.Martin, F. Sonnenburg	Managerial Ownership Changes and Mutual Fund Performance
16-02	A.Gargano, A. G. Rossi, R. Wermers	The Freedom of Information Act and the Race Towards Information Acquisition
16-01	G. Cici, S. Gibson, C. Rosenfeld	Cross-Company Effects of Common Ownership: Dealings Between Borrowers and Lenders With a Common Blockholder

This document only covers the most recent CFR working papers. A full list can be found at www.cfr-cologne.de.

centre for financial research
cfr/university of cologne
Albertus-Magnus-Platz
D-50923 cologne
Fon +49(0)221-470-6995
Fax +49(0)221-470-3992
kempf@cfr-cologne.de
www.cfr-cologne.de