Path-dependent Risk Measures

– Theory and Applications –

Dissertation

submitted to the Faculty of Business and Economics
at the Georg-August-Universität Göttingen

by

Philipp Maximilian Möller, M.Sc.

from Hannover

Göttingen, November 18, 2020

Thesis Committee:

Prof. Dr. Olaf Korn

Prof. Dr. Stefan Dierkes

Jun.-Prof. Dr. Holger A. Rau

Contents

List of Figures

List of Tables

Introduction

Taking heed of risk is a pivot of finance. It would be almost absurd imagining Markowitz portfolio theory without sigma, the Capital Asset Pricing Model without beta, Sharpe ratios without the standard deviation, banks without market, credit, or liquidity risk, or option pricing with Brownian motion without motion. Although risk is deeply ingrained in finance, there is no consensus about its measurement. Some intuitive and tractable measures have taken center stage, for example, the (conditional) value-at-risk (VaR) is the standard risk measure for setting capital requirements for trading portfolios (Alexander and Baptista, 2017). Most current risk measures have in common that in quantifying risk for a specific period they only consider the risk at the end of the period. For example, the 10-day VaR indicates the cumulative loss at day 10 that is not exceeded with a prescribed probability. Likewise, the conditional VaR, all lower partial moments, and the standard deviation are functions of the final return distribution.[1] Of course, the final outcome matters – but often so does the path. As anyone stuck in a traffic jam can tell, it also matters *how* you get there. To account for this (fairly obvious) insight, path-dependent risk measures have been developed, which also take into account the path that leads to the final investment outcome. Path dependence matters in finance for several reasons. First, an investment's intermediate price movements may cause liquidity issues, for example, when margin calls force investors to liquidate a position prematurely. Second, the price path may be highly relevant for psychological reasons. An investment that steadily increases in value is likely to be perceived differently than an investment with volatile price movements. From a fund manager's perspective, both positive and negative intermediate price movements can cause a withdrawal of funds – either because of cashing in gains or because of losing trust in the manager.[2]

---

[1]While the standard deviation as a risk measure requires only the distribution of returns at the end of the period, its most common estimator is derived from higher-frequency returns and square-root-of-time scaling. Therefore, the estimator requires higher-frequency returns but still disregards their order.

[2]Although negative price movements may seem more relevant, it is well-documented empirically that investors are often quicker to realize gains than losses – a phenomenon coined *disposition effect* by Shefrin and Statman (1985).

Among the most widely quoted path-dependent risk measures are drawdown measures (Mahmoud, 2017). They quantify risk by considering the relative loss incurred from a previous peak. For example, an investment that initially costs \$ 40, moves up to \$ 50 and then falls back to \$ 40 incurs a drawdown of 0.2 because it has lost 20% relative to the running maximum at \$ 50. Constructed in terms of the running maximum, all drawdown measures are inherently path-dependent. The running maximum serves as a natural reference point because it definitely exists, it is easy to find, and it is straightforward to interpret. At best, the drawdown at a given point in time is zero, which occurs when the price resides at an all-time high. In all other cases, investors face a loss with respect to the running maximum. If such a loss is realized at end of the investment horizon, investors are likely to experience regret for not having exited the investment at the maximum. In the above example, the final drawdown of 20% thus expresses the loss compared to selling the investment with optimal market timing.[3] By strictly focusing on losses, all drawdown measures are asymmetric risk measures.

The primary application of drawdown measures is in asset management, where they are used to quote performance, to allocate or redeem funds, or to evaluate fund managers (Landriault et al., 2015; van Hemert et al., 2020). Drawdown measures are typically used around hedge funds, funds of hedge funds, or mutual funds, but also stock or commodity investments (see, e.g., Eling and Schuhmacher (2007), Heidorn et al. (2009), Eling (2008), Kim (2018), Auer (2015), respectively). While the Commodity Futures Trading Commission mandates commodity trading advisors to report their drawdown[4], hedge fund managers usually voluntarily disclose this information (Lhabitant, 2004). Garcia and Gould (1987) already noted over 30 years ago that, in their experience, despite the variance being accepted as a "good measure of risk," many investors put more emphasis on drawdown. Burghardt and Walls (2011, p. xiii) note that the inspiration for their work on drawdown "was a rough survey we took at one of our earliest conferences. We had asked everyone what they thought the most useful measure of risk was, and a very large majority replied 'drawdown.'"

Consequentially, academic and practical literature about drawdown has developed, a comprehensive review of which can be found in Section 3.2. Generally, several broad strands of literature have emerged. By formulating a continuous-time investment problem with a drawdown constraint, Grossman and Zhou (1993) initiate a strand of literature regarding drawdown constraints. These

---

[3]This notion requires assuming zero interest rates.

[4]See section 17 CFR §4.35 on performance disclosures.

constraints are intended to ensure that an investment's value never falls below a fixed percentage of the running maximum at any time. Subsequent work on investment problems with a drawdown constraint includes Cvitanic and Karatzas (1995), Alexander and Baptista (2006), Elie and Touzi (2008), Sekine (2013), Yao et al. (2013), Cherny and Obłój (2013), Rieder and Wittlinger (2014), Angoshtari et al. (2016), Kardaras et al. (2017), and Roche (2019). A second strand of literature addresses mathematical properties of the drawdown process. Given an underlying stochastic process, the drawdown process is usually defined as the absolute difference between the underlying's current value and its maximum. Under various assumptions, several properties of the drawdown process have been investigated, which are often related to stopping times, see Hadjiliadis and Vecer (2006), Mijatović and Pistorius (2012), Landriault et al. (2017b), and Bai and Liu (2019). Drawdown measures have also been used in the denominator of performance ratios, and a different strand of literature has specifically addressed this application. What is debated is whether performance ratios based on different (drawdown) measures lead to different rankings of investments. For various sets of drawdown measures and various types of investments, this question is addressed in Eling and Schuhmacher (2007), Eling (2008), Caporin and Lisi (2011), Haas Ornelas et al. (2012), Auer and Schuhmacher (2013), and Auer (2015). Although all of these strands focus on drawdown – either on drawdown constraints, drawdown processes, or drawdown-based performance ratios – they do not specifically address drawdown risk measures. Regarding risk measures, Chekhlov et al. (2005) seminally define the conditional drawdown (CDD), which constitutes a family of drawdown risk measures. It includes the maximum drawdown (MDD) and the average drawdown (ADD), which are arguably the most prominent drawdown risk measures to date. An additional family, the conditional expected drawdown (CED), has been defined recently in Goldberg and Mahmoud (2017). It is theoretically interesting but less applicable in practice because it requires knowledge of tail means of the MDD distribution. Properties of drawdown measures are almost exclusively studied for the MDD. Its distribution is analyzed by Magdon-Ismail et al. (2004), Cheridito et al. (2012), and Casati and Tabachnik (2013). Comparative statics are computed in van Hemert et al. (2020) to analyze how changes in return, volatility, length of time horizon, and autocorrelation affect the MDD.[5] Despite the aforementioned literature, drawdown measures remain much less well-studied compared to more conventional risk measures, such as value-at-risk and volatility (Goldberg

---

[5]The work of van Hemert et al. (2020) probably comes closest to this dissertation's objective of analyzing properties of drawdown measures. It partly follows the approach used in the first paper of this dissertation, but there are notable differences. For example, their comparative statics do not control for higher moments, they use much coarser data, and they only consider the MDD.

and Mahmoud, 2017). Especially when it comes to properties of drawdown measures, academic literature is almost non-existent.

It is this void that this dissertation intends to address. Due to a lack of previous research, hardly anything is known about fundamental aspects of drawdown measures. Are drawdown measures sensitive to an asset's return, its volatility, or both? Do assets with skewed or particularly fat-tailed returns have significantly higher drawdown? Does it matter which drawdown measure is used to assess an investment, or do they all come to the same conclusion? Are future drawdowns in any way predictable given drawdowns measured today? This dissertation addresses all of these questions in three major chapters, each of which takes the form of a paper and can be read independently of each other. Their objectives, methods, and outcomes summarize as follows.

The first paper *Drawdown Measures and Return Moments* addresses fundamental properties of drawdown measures, most notably their relation to return moments. As drawdown measures are computed from asset paths, the central idea is to investigate how the drawdown changes when properties of the path change. The first four moments of the investment's returns – i.e., its expected value, standard deviation, skewness, and kurtosis – are a natural set of properties for the first investigation of this kind regarding drawdown measures. These statistical moments are standard ways to characterize assets as trending upward or downward, being stable or volatile, being symmetric or asymmetric (i.e., expressing left-, right-, or zero skewness), or having fat tails. In order to investigate the moment effects most precisely, a simulation framework is suited best. Ideally, each moment is varied independently of the others to isolate its effect. A stochastic process that satisfies this requirement remarkably well is the exponential Lévy process with normal inverse Gaussian increments. Its parameters can be steered such that paths with realistic combinations of the first four return moments can be simulated. For the resulting paths, the MDD, ADD, and $CDD_{0.8}$ from the conditional drawdown family are computed; the conditional expected drawdown is computed for four confidence levels between 0.80 and 0.95. Additionally, a new end-of-period drawdown (eopDD) is postulated as the return difference between the ex-post best strategy with optimal market timing and the actually experienced path. For these drawdown quantities, comparative statics are simulated, i.e., starting from a base case, each return moment is varied separately while holding the other three moments constant. Unanimously, higher first moments cause lower drawdown, and higher second moments cause higher drawdown. Both results are plausible: Higher expected returns lead to fewer losses, thus also curbing drawdown; higher volatility leads to both up and down movements, thus incurring losses and higher drawdown. For skewness and kurtosis, the effects differ between drawdown

measures. With increasing right-skewness, all CED measures and the MDD decline considerably, but $CDD_{0.80}$ declines only slightly, while eopDD and ADD remain constant. Increasingly fat tails cause a decline in all CDD measures and eopDD, while the CED increases or declines slightly depending on the confidence level. As most people prefer investments with high odd moments and low even moments, the CED with high confidence levels captures this intuition most successfully.

Several smaller analyses complement the main findings. First, a tailor-made jump diffusion model with a discrete jump height distribution is employed to reproduce the previous results with a clearly different model. Second, both models are modified to investigate the effect of sudden random shocks on the drawdown measures. Upward shocks do not affect the drawdown measures; downward shocks substantially increase drawdown whereas the effect scales approximately linearly with the shocks' magnitude. Third, the effects of dependencies in the return time series are analyzed in an autoregressive (AR) and a generalized autoregressive conditional heteroscedasticity (GARCH) model. Holding the first two moments constant, autocorrelation in the return variance does not affect drawdown in the GARCH(1,1) model. The same holds for autocorrelation in the returns, which seemingly contradicts results from Goldberg and Mahmoud (2017) for the AR(1) model. Further analysis shows that the results in the literature had imprecisely been interpreted as an effect of autocorrelation while instead being an effect of variance. As a last additional analysis, the new eopDD measure is compared to the other drawdown measures empirically. To this end, random portfolios of S&P 500 stocks are assembled, and the drawdown of their paths is compared using rank correlations. The correlation between eopDD and the other measures is positive but considerably lower than the correlation between, for example, maximum and average drawdown, indicating that the new eopDD measure captures different aspects of drawdown.

In the second paper[6] *Drawdown Measures: Are They All the Same?*, the idea of assessing differences and similarities between drawdown measures is developed further. This endeavor is motivated by an ever-growing variety of drawdown measures introduced in the literature. Some similarity between drawdown measures is, of course, to be expected because all drawdown measures follow similar principles, such as considering losses from a previous peak. However, too much similarity would indicate that defining ever more complex drawdown measures is not worthwhile, and that a focus on a few promising measures may streamline further research. In order to compare drawdown measures on theoretical grounds, a first attempt may be to

---

[6]This paper is co-authored by Olaf Korn and Christian Schwehm.

check theoretical properties like homogeneity or subadditivity. However, such an analysis is not particularly fruitful because many drawdown measures satisfy the conditions of being a generalized deviation measure according to Rockafellar et al. (2006) in a similar fashion, not revealing notable differences. Nonetheless, an alternative approach is developed in the paper. The majority of drawdown measures from the literature can be decomposed into a drawdown graph and a weight function; the drawdown graph depends on the asset path but is identical for all drawdown measures, while the weight function is independent of the asset path but characteristic of the drawdown measure. Comparing the weight functions provides an intuitive grasp of which aspect of drawdown each measure emphasizes because similarities between weight functions point to similarities between drawdown measures. For example, the difference between MDD and ADD becomes quite apparent in the weight functions as the MDD's weight function is zero everywhere except for a single element of the drawdown graph, while the ADD's weight function is constant, thus distributing weight equally along the drawdown graph.

Complementing this approach, drawdown measures are also compared empirically. To this end, the idea of computing rank correlations for random portfolios is adopted from the first paper. Here, however, the construction of portfolios is more sophisticated: 1,000 fictitious portfolio managers randomly pick stocks from the MSCI World universe such that they hold 100 stocks at any time. Country and sector constraints ensure that the proportion of stocks from any given country or sector does not deviate drastically from the corresponding proportions in the MSCI World index. Each month, there is some rebalancing and adjustment for stocks leaving the index, and each new stock receives a weight between 0 and 2%. In the random setup, all fictitious managers have a hit ratio of 50%, i.e., their probability of picking future winners that have above median return is 0.5. In hindsight, however, some portfolio managers can be endowed with higher hit ratios by providing them with a higher likelihood of identifying future winners over future losers. Therefore, each time the skillful portfolio managers create a portfolio and add or drop stocks during rebalancing, the odds of picking a winning stock and dropping a losing stock are in their favor. Once 1,000 portfolio paths are simulated with and without skill, the portfolio managers are ranked using different drawdown measures, and these ranks are compared via correlation coefficients. Both for hit ratios of 0.5 and 0.6, the results are similar: All correlations are positive, ranging between 0.258 and 0.874. Thus, correlations between drawdown measures differ substantially. Especially with an eye toward eopDD, with which correlations are lowest, drawdown measures do not appear to be "all the same".

When some managers have stock picking skill – because they have above average hit ratios –

and other managers do not, the question arises whether drawdown measures can help identify these skillful managers. For example, if samples of 1,000 skilled managers with a hit ratio of 0.6 and 1,000 unskilled managers with a hit ratio of 0.5 are merged, could drawdown measures tell them apart? Naturally, the skilled managers' portfolios should incur lower drawdowns. Hence, if among the 1,000 managers with lowest drawdown all were skilled, the drawdown measure would perfectly identify skill. Using the ratio of skilled managers out of 1,000 as a measure of skill detection, all drawdown measures can indeed detect skill. The ADD correctly identifies approximately 75% of managers, the MDD slightly more than 65%, and the eopDD slightly less than 65%. All drawdown measures outperform the standard deviation and expected shortfall, which are used for comparison.

As mentioned above, an entire strand of drawdown literature is concerned with drawdown-based performance ratios. Do these ratios behave differently when it comes to skill and rank correlations? Regarding the latter, different drawdown ratios rank portfolios quite similarly, with correlations ranging between 0.455 and 0.937. Regarding skill detection, all drawdown ratios (except for the ratio based on eopDD) perform well and virtually the same. However, all drawdown ratios sometimes fail miserably at identifying skill in periods when the return is negative because the negative return affects both numerator and denominator of the ratio and can lead to unskilled managers receiving better drawdown ratios. In summary, drawdown ratios are more similar than drawdown measures and prove to be problematic in detecting skill in some circumstances.

The third paper *Drawdown Persistence and a Convenient Shortcut to Predicting Mutual Fund Drawdown* focuses on persistence. Since this aspect of risk measures has not been addressed for drawdown measures before, it is unclear whether the drawdown risk identified for an asset ex-ante is indicative of this asset's drawdown in the future, i.e., whether an asset's drawdown risk persists. For any risk measure, persistence is a significant property because information about the past is then potentially valuable for the future. For example, funds regularly report their historical drawdown to showcase management success (Lhabitant, 2004); however, without persistence, the fund managers' drawdown track record would bear no information about their future drawdown whatsoever, and, thus, reporting drawdown would be pointless in the first place. In a sample of more than 7,000 equity mutual funds, significant relative persistence can be documented for the maximum and average drawdown. Thus, on average, funds with lower drawdown than their peers in the past continue to have lower drawdown in the future. Methodologically, persistence is assessed with two independent measurement approaches: the correlation between the asset rankings of the two periods, and the ratio of the average future drawdown of the funds in the

highest historical drawdown decile divided by the drawdown of the funds in the lowest decile. For both measures, strong persistence is evidenced in the full sample as well as in subsamples regarding fund types and period lengths.

A question that is closely related to persistence is that of prediction. While persistence is concerned with how historical drawdown information relates to the future drawdown, prediction allows for *any* historical information to be used to infer future drawdown. Because exhaustively searching for the best prediction is tedious and likely incomplete, a viable option is to test several promising predictors. From these, historical volatility emerges as a predictor that outperforms even the drawdown measures themselves, i.e., the historical standard deviation contains more helpful information about future MDD than the past MDD. This does not seem to be incidental because it replicates for all fund categories. A closer look at the data reveals interesting relations: Within the same year, the return is strongly negatively correlated with the MDD, while the standard deviation is even more strongly positively correlated with the MDD. However, only the standard deviation persists over time whereas the return does not. From these observations, two hypotheses may be formulated: First, volatility persistence combined with the same-period correlation between standard deviation (SD) and MDD is the driver behind MDD persistence and SD-MDD persistence. Second, if the return is persistent, so is the MDD. What is required to assess these hypotheses is a simulation model that incorporates both return and volatility persistence. Such a model – constructed with fractionally integrated and independent mean and variance processes – indeed supports both hypotheses. Therefore, whether it is best to predict drawdown with past volatility or past drawdown depends on the persistence of return and volatility in the data.

After summarizing the three main chapters separately, several overarching aspects shall be discussed next, such as the definition of drawdown measures. Regarding its literal meaning, "drawdown" signifies a process of depletion or reduction. It is a common term in hydrology, used to describe the depletion process of water reservoirs or aquifers (Wu et al., 2016). In finance, drawdown refers to a realized or unrealized financial loss, usually large in size. The term is often used without an explicit definition as if assuming some tacit agreement what is meant, for example, in Eling (2008) or Tashman and Frey (2009). Judging from the diverse definitions present in the literature, no such agreement actually exists. For example, some sources count only such losses as drawdowns that have not been interrupted by an uptick in prices (e.g., Schuhmacher and Eling (2011)). A new series of losses (and, thus, a new drawdown) starts as soon as a period with positive returns interrupts a series of negative returns. This continuous

drawdown measure is, of course, highly unreliable because a tiny difference in the path can cut the drawdown in half – depending on whether a return is slightly above or below zero. A second example of differing definitions is concerned with the measurement of losses. Usually, losses from the previous peak are quantified *relatively* by using returns; however, some non-finance articles use the *absolute* difference between the maximum and the current value. The latter is mathematically more tractable but much less applicable in practice since the drawdown of investments with different initial values cannot be compared directly. When drawdown measures are defined differently, any conclusions must be considered with caution because it can only be speculated how results for one drawdown definition transfer to another.

Therefore, in this dissertation, extra emphasis is placed on providing thorough and logically consistent definitions of all drawdown measures. All such definitions start with the drawdown graph. Given a discretized price path,[7] it is defined for each point in time t as the percentage loss from the running maximum until t. Figure 0.1 depicts an exemplary price path and its corresponding drawdown graph.
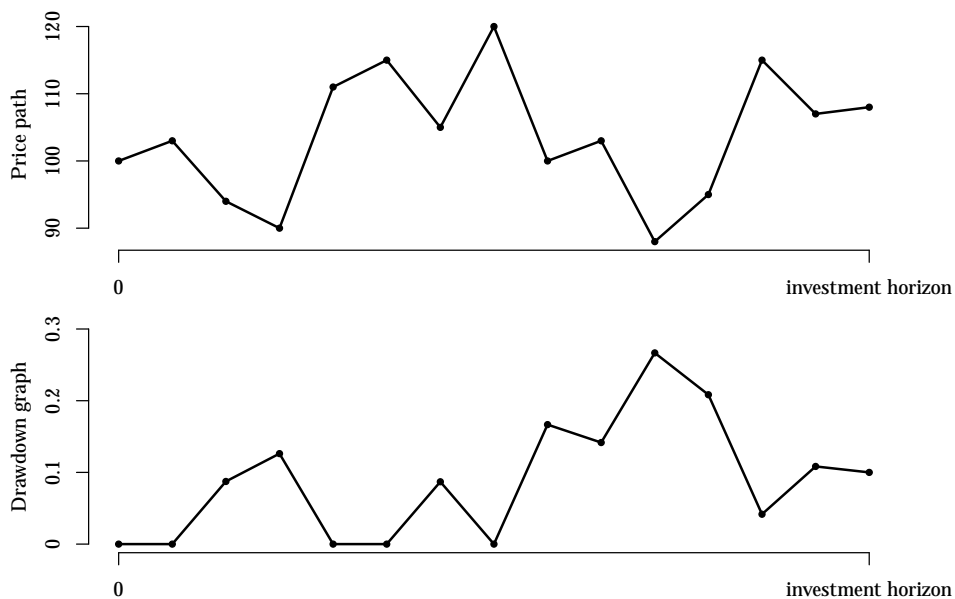
Figure 0.1: Illustration of an asset's discrete price path (above) and the corresponding drawdown graph (below). Defined as the relative loss between the price at the current position and the preceding all-time high, the drawdown graph is an immediate consequence of the price path.

By definition, the drawdown graph starts at zero, and always returns to zero whenever the asset graph reaches a new all-time high. It cannot fall below zero because any point that obtains a

---

[7]Drawdown definitions in a continuous-time setting are discussed in Chapter 1.

negative loss (i.e., a gain) with respect to the running maximum constitutes a new maximum, and the loss with respect to it is again zero. Sometimes, the drawdown graph resembles a mirror image of the price path because a decline in the price path corresponds to an increase in the drawdown graph. However, the mirror image can be far from perfect; for example, when the price path climbs from one running maximum to another, the drawdown graph remains flat at zero. Naturally, the maximum drawdown is the maximum and the average drawdown is the average of the drawdown graph. Considering a specific investment horizon, the end-of-period drawdown introduced in Chapter 1 is also intuitive as it coincides with the last point of the drawdown graph, and emphasizes the regret experienced at the end of the investment horizon when looking back at the asset's path. It reflects that it can make a difference if an asset is in deep drawdown when it is evaluated, or if an early downturn has been recovered in the meantime. All these risk measures are generalized deviation measures according to Rockafellar et al. (2006). Additionally, the drawdown graph is the basis for the largest class of drawdown measures proposed in the literature so far – the weighted drawdown (wDD) framework, introduced in Chapter 2. In this framework, a weight function assigns weights to each element of the drawdown graph and the drawdown measure results as a weighted sum. This way, each admissible weight function corresponds to a unique drawdown measure. Not only can almost all drawdown measures proposed in the literature be subsumed under the wDD framework, but new drawdown measures can be defined with ease. For example, a linearly increasing weight function may incorporate the idea that drawdowns become more painful toward the end of the investment horizon. A trend-dependent weight function may allow drawdowns during a downturn to receive higher weights because they are potentially more painful than remaining drawdown during a recovery. Both this linearly weighted drawdown (lwDD) and the trend weighted drawdown (twDD) are investigated in Chapter 2. The flexibility of the wDD framework provides investors with an easy tool to construct a drawdown risk measure that suits their risk preferences.

Different sets of drawdown measures are used in the three main chapters on purpose. The objective of the first paper is to assess how changes in return moments and autocorrelation affect drawdown measures in general. Therefore, a broad spectrum of drawdown measures is desirable, which is why all "standard" drawdown measures from the literature are used. Apart from MDD and ADD, which are the two extremes of the CDD family, the intermediate measure $CDD_{0.8}$ from this family is also included. Four members of the CED family, $CED_{0.8}$, $CED_{0.85}$, $CED_{0.9}$, $CED_{0.95}$, are investigated, as is the new eopDD. Analyzing a broad variety of drawdown measures turns out to be beneficial because the effects of higher moments differ between drawdown

measures. In the second paper, a key aspect is the comparison of different drawdown measures. Therefore, the set of drawdown measures is augmented to include the average squared drawdown as well as the lwDD and the twDD from the new weighted drawdown framework. Because it is infeasible for this type of analysis, the CED is omitted from the second paper; because it is very similar to the MDD, the $CDD_{0.8}$ is omitted as well. Additionally, all drawdown measures are used in the denominator of the corresponding performance ratios. The third paper focuses on identifying first evidence of drawdown persistence. Therefore, the arguably most wide-spread drawdown measures, MDD and ADD, are employed first. As drawdown persistence is present, and there is little reason to expect that other drawdown measures yield drastically different results, there is no immediate need to extend the set of drawdown measures.

Analytic results for drawdown-related quantities are exceedingly hard to come by. Due to the drawdown measures' path dependence, it does not suffice to make an assumption about a return distribution. Instead, stochastic processes have to be assumed to obtain a path from which a drawdown measure can subsequently be derived. Stochastic processes significantly exceed probability distributions in complexity because additional assumptions about the dependence structure, the time-homogeneity of the probability distribution, and the continuity of the path become necessary. For example, assuming a standard Brownian motion, Magdon-Ismail et al. (2004) succeed in deriving a closed-form solution for the expectation of maximum drawdown only when the drift of the Brownian motion is zero. For positive or negative drift, the solution involves an infinite sum in which each summand contains an eigenvalue problem. To complicate matters further, we focus on the economically relevant definition of drawdown in terms of relative returns, but all analytic results are obtained defining drawdown in absolute terms. Therefore, simulation studies and empirical analyses are pursued as viable alternatives, each modeling choice depending on the respective research question. Since moment properties are vital in the first paper, normal inverse Gaussian Lévy processes and a specifically designed jump-diffusion model are used to generate sufficiently non-normal returns. Both models satisfy the challenging requirement that the first four return moments can be varied almost separately. Moreover, both models are structurally different with one having infinite activity jumps and semi-heavy tails and the other having no heavy tails and few, fixed-height jumps. As the dependence structure is not essential, both processes have independent increments, which simplifies the simulations. In the explicit analysis of autocorrelation in returns and volatility, AR and GARCH processes are employed. As the first two moments are highly influential and could easily skew the results, extra emphasis is placed on varying only the autocorrelation but not the first two moments

such that the specific effects of autocorrelation can be captured. Both AR(1) and GARCH(1,1) model short-term autocorrelation; conversely, the fractionally integrated (FI) processes in the third paper also model long-term autocorrelation, which is necessary because the third paper is concerned with persistence. To investigate both return and volatility persistence separately, a stochastic mean stochastic variance model is employed whose components are long-range dependent but independent of each other. As drawdown measures are applied mostly in finance and it is rarely possible to model each and every aspect of real data, empirical analyses nicely complement the simulated findings. In the third paper, for example, where mutual funds are used to investigate into relative drawdown persistence, some heterogeneity in the cross section is explicitly necessary. In the second paper, random portfolios are simulated based on the MSCI World constituents and their respective return time series. This mixture of stock data and simulations has two reasons: First, manager skill can be modeled explicitly, which facilitates an investigation into whether drawdown measures can detect skill. Second, realistic time series are obtained for the comparison of different drawdown measures. By simulating portfolios from a common universe, the managers' portfolio paths are neither too similar nor too different. If the paths were too similar, even the most potent risk measures could not detect meaningful differences and all rankings would result from chance; if the paths were too different, any risk measure could tell the difference and little would be learned. Overall, this dissertation addresses questions about drawdown resorting to simulations, empirical analysis, and a mixture of both. All analytic results are not about properties of drawdown measures but usually concern properties of the stochastic processes used to conduct the simulation studies.

First and foremost, the papers in this dissertation intend to cater to an audience in academia as all chapters seek to advance the literature on drawdown-based risk measures from a different angle. To the best of the author's knowledge, it is the first dissertation with this objective. Apart from results on drawdown properties, this work supplies detailed definitions of drawdown measures, a new drawdown framework to provide structure, and comprehensive references of the status quo of drawdown research. As such, it should constitute an ideal basis for future research projects regarding drawdown. Since the drawdown literature appears to trail the application of drawdown measures in investment practice, this dissertation may also be of interest to finance practitioners. It may theoretically support ideas previously used without rigorous evidence, or help set straight incorrect beliefs. As drawdown measures are versatile and often counter-intuitive it also aspires to be an entertaining and thought-provoking read.

1 Drawdown Measures and Return Moments

Published in the International Journal of Theoretical and Applied Finance.[1]

Abstract

This paper provides an investigation of the effects of an investment's return moments on drawdown-based measures of risk, including Maximum Drawdown (MDD), Conditional Drawdown (CDD), and Conditional Expected Drawdown (CED). Additionally, a new end-of-period drawdown measure is introduced, which incorporates a psychological aspect of risk perception that previous drawdown measures had been unable to capture. While simulation results indicate many similarities in the first and second moments, skewness and kurtosis affect different drawdown measures in radically different ways. Thus, users should assess whether their choice of drawdown measure accurately reflects the kind of risk they want to measure.

[1]This chapter is a version of an article published in the International Journal of Theoretical and Applied Finance, Vol. 21, No. 7, 1850042 (2018), https://doi.org/10.1142/S0219024918500425. The copyright is owned by World Scientific Publishing Company, https://www.worldscientific.com/worldscinet/ijtaf. Reprinted with kind permission.

1.1 Introduction

In managing risks, the choice of a suitable risk measure is vital. Many classical risk measures like (semi-)variance, lower partial moments, VaR, and expected shortfall are functions of the distribution of the risky object at the end of the investment horizon alone. There are two reasons why instead a path-dependent concept may be necessary. First, conditions on the path may ensure a strategy's feasibility, for example, if liquidity constraints require the path not to fall below some threshold level. Some funds may even face automatic shutdown if a drawdown constraint is breached during the investment period (Chekhlov et al., 2005). Second, the path may be an essential psychological component: Investors will most likely perceive and evaluate two investments with the same 5% final return and equal standard deviation differently if one strategy had gained 20% but the other had lost 20% in the meantime. Differently perceived risks can have substantial consequences for fund managers if investors choose to withdraw funding. Since drawdown measures are frequently used to report a fund's performance, managers with a poor drawdown history may also find it difficult to attract new capital. Thus, fund managers have a strong incentive to construct portfolios that avoid drawdowns.

In the first studies on drawdown, the constraint aspect was dominant, i.e., portfolio optimization was pursued under a drawdown constraint, for example, in Grossman and Zhou (1993) and Cvitanic and Karatzas (1995). Since then, two classes of drawdown measures have been introduced: the Conditional Drawdown (CDD) family introduced by Chekhlov et al. (2005) and the Conditional Expected Drawdown (CED) family introduced by Goldberg and Mahmoud (2017). The CDD family includes the Maximum Drawdown, i.e., the single largest peak to valley loss, which is commonly applied in practice. The CED family builds on the distribution of Maximum Drawdowns. Beyond the constraint aspect, the psychological aspect of drawdown can be incorporated by emphasizing the time horizon. Often, practitioners encounter fixed investment horizons, evaluation periods, or portfolio rebalancing cycles where the performance at the end of the time horizon is key. For the aforementioned risk measures, it is, however, irrelevant if the negative turn causing, for example, a large Maximum Drawdown is at the beginning of the investment period or towards its end. Hence, these risk measures cannot capture the psychological component. Therefore, we introduce a new end-of-period drawdown measure (eopDD), which focuses on the drawdown at the end of the investment horizon, and analyze its properties. Empirically, portfolio rankings by the eopDD differ markedly from rankings by other drawdown measures.

Drawdown risk measures have attracted considerable attention in portfolio management and academia in the last two decades. Comprehensive references are available in Mahmoud (2017) and Schuhmacher and Eling (2011). Very recent developments include an intensified probe into mathematical properties (Landriault et al., 2017b) as well as drawdown measures carving their way into fields like control theory (Hsieh and Barmish, 2017), option pricing (Dassios and Lim, 2018), insurance (Palmowski and Tumilewicz, 2017), and energy markets (Charwand et al., 2017). Despite ongoing research efforts, fundamental properties of drawdown-based measures are not yet understood. For example, it is an open question how skewness and kurtosis of the underlying asset returns affect drawdown measures. The presence of skewness and kurtosis in asset returns has been documented extensively in the literature, for example, in Adcock et al. (2015) and Mandelbrot (1963). While the importance of considering higher moments is frequently emphasized in the finance literature (Dittmar, 2002; Guidolin and Timmermann, 2008; Maringer and Parpas, 2009; Harvey et al., 2010; Jondeau and Rockinger, 2012), this paper is the first to relate higher moments to drawdown risk measures. It offers a detailed account of how an investment's return moments influence drawdown-based measures, including Maximum Drawdown, Conditional Drawdown, Conditional Expected Drawdown, and the new end-of-period drawdown.

To provide meaningful guidance for decision-making under risk, risk measures should adequately reflect how investors perceive risk. There is both theoretical and empirical evidence that high odd moments are considered desirable and high even moments are not. Intuitively, investors appreciate high returns (first moment), dislike high variance (second moment), favor extreme positive over extreme negative events (third moment), and dislike exposure to heavy tails (fourth moment). Consistent with this intuition, Scott and Horvath (1980) prove that in a theoretical setting there should be a preference for high skewness and a preference for low kurtosis. Harvey and Siddique (2000) find that investors demand a significant premium for stocks with negative skewness. Dittmar (2002) shows that incorporating an additional aversion to kurtosis further improves the explanation of the cross section of expected returns. Proceeding from these results, this paper addresses the question which drawdown measure captures moment preferences most accurately.

Apart from moment influences, we investigate the effect of upward and downward shocks on drawdown measures. To this end, we introduce randomly occurring market shocks of varying magnitude. This is especially interesting for path-dependent risk measures since large downward jumps at the end of the time horizon are likely to facilitate a large end-of-period drawdown

whereas a large downward jump at the beginning may be made up over the course of the path. Hence, it is unclear what aggregate effect jumps have on average if their timing is random.

Drawdown measures are path-dependent by nature. Therefore, we propose a simulation framework to simulate entire sample paths with the desired return moments. The exponential Lévy process with normal inverse Gaussian increments as introduced by Barndorff-Nielsen (1997) best serves our needs because it allows modeling higher moments like asymmetry and fat tails in a realistic way. Simulating paths with certain moment properties allows us to examine the effect of varying return moments on a wide range of drawdown measures. In detail, we apply the model to the Maximum Drawdown, Average Drawdown, and $CDD_{0.80}$ from the Conditional Drawdown family, the Conditional Expected Drawdown CED_α for several confidence levels α, and the new end-of-period drawdown including a decomposition into the probability of drawdown and the conditional expected end-of-period drawdown. For robustness, we repeat the simulation study for a second model where we add a straightforward jump mechanism to geometric Brownian motion to obtain a jump diffusion process. For both models, we explicitly derive the choice of process parameters which are necessary to generate the first four moments of annual returns.

For the first and second moment, the simulation study validates common intuition. Strategies with higher expected return have lower drawdown because of the upward trend. If a strategy's standard deviation increases, its drawdown increases as well. While these results are valid for all drawdown-based measures under consideration, the results for higher moments differ substantially. Intuitively, negative skewness should c. p. induce more drawdown risk and thus lead to higher drawdown risk measures. However, only the CED is consistent with this intuition. The CDD exhibits non-monotonous behavior with the Average Drawdown being almost unaffected by changes in skewness. Higher kurtosis counter-intuitively yields lower drawdown measures almost across the board. Only for the CED with high confidence levels α, drawdown increases with kurtosis. The size of the kurtosis effect is strong for the CDD but almost negligible for the CED. For all moments, the directions of the new end-of-period drawdown almost perfectly resemble the directions of the Average Drawdown. In total, the CED with high α is most in line with standard moment preferences.

To pin down the moment effects, each moment is varied ceteris paribus. In contrast, we also model the effect of shocks within a different setup where many moments are impacted simultaneously. We find that the direction matters: Downward shocks strongly increase drawdown measures unanimously whereas upward shocks have no significant influence. The size of the downward

jumps translates into increases in drawdown approximately linearly.

Finally, we analyze what influence autocorrelation has on the drawdown measures. As both momentum and volatility clustering have been found in returns, we consider autocorrelation both in the mean and variance equations in an AR and GARCH model, respectively. For all drawdown measures, the effects of autocorrelation are almost negligibly small. However, one has to be careful to correct for mean and especially variance because otherwise autocorrelation drives up variance which, in turn, increases the drawdown measures. These findings explain and significantly extend the results on autocorrelation in Goldberg and Mahmoud (2017).

The rest of this paper is organized as follows. In section 1.2, we provide definitions of all standard drawdown risk measures. Afterwards, we motivate and introduce the new end-of-period drawdown measure and analyze its properties. In section 1.3, the simulation framework is introduced and illustrated. Section 1.4 contains the results of the simulation study and answers the research questions regarding the behavior of the drawdown measures, while section 1.5 discusses robustness. In section 1.6 and 1.7, we analyze the effects of additional jumps and autocorrelation, respectively. Section 1.8 concludes.

1.2 Drawdown-Based Risk Measures

Several drawdown-based measures have been proposed in the literature. For the sake of completeness, we first provide all definitions of the standard drawdown measures. Afterwards, a new end-of-period drawdown measure is introduced, and its properties are examined.

1.2.1 Setting

We fix a time interval $[0, T]$ where T may be the investment horizon of an investor or the time of performance evaluation of a portfolio manager. Consider a strategy[2] $\mathbf{S} = \{\mathbf{S}_t\}_{t \in [0,T]}$, for example, the price process of a single asset or of a portfolio, and its running maximum $\mathbf{M}_t = \sup_{u \in [0,t]} \mathbf{S}_u$ for which we require $\mathbf{M}_t > 0, t \in [0, T]$. The drawdown curve of the strategy is then given by $\mathbf{D}_t = \frac{\mathbf{M}_t - \mathbf{S}_t}{\mathbf{M}_t}$, which at each point t indicates the relative loss from the running maximum until time t. In portfolio management practice, this curve is known as the underwater curve (Zabarankin et al., 2014). All drawdown-based measures are functionals of this drawdown curve.

---

[2]The notation is deliberately general such that it is applicable to arbitrary strategies as long as the drawdown functionals mentioned below exist.

In all applications, we consider discretizations of the continuous process $\mathbf{S}$ with $N+1$ equidistant points, and denote this sequence of random variables by $S = (S_0, \ldots, S_N)$ where $S_0 = \mathbf{S}_0$ and $S_N = \mathbf{S}_T$. Denote by M_i the maximum of the first i instances of S with $i = 1, \ldots, N$. The discrete time analogue to the drawdown curve is the drawdown vector $D_i = \frac{M_i - S_i}{M_i}$. In the following, we define drawdown measures both for the continuous time process $\mathbf{S}$ and for the discretized process S.

1.2.2 Standard drawdown measures

First, the Maximum Drawdown (MDD) is defined as the supremum of the drawdown curve $\text{MDD}(\mathbf{S}) = \sup_{t \in [0,T]} \mathbf{D}_t$, i.e., the maximal loss incurred from peak to trough. For the discrete process, the MDD is the maximum of the drawdown vector $\text{MDD}(S) = \max_{1 \leq i \leq N} D_i$. The Average Drawdown (ADD) is defined as the standardized integral under the drawdown curve $\text{ADD}(\mathbf{S}) = \frac{1}{T} \int_0^T \mathbf{D}_t \, \mathrm{d}t$ or for the discretized process as the averaged drawdown vector $\text{ADD}(S) = \frac{1}{N} \sum_{i=1}^{N} D_i$. Both MDD and ADD belong to the family of Conditional Drawdown (CDD) measures introduced by Chekhlov et al. (2005). The CDD with confidence level α is loosely defined as the mean of the worst $(1 - \alpha) \cdot 100\%$ drawdowns. In detail,

$$\text{CDD}_\alpha(\mathbf{S}) = \left(\int_0^T \mathbb{1}\{\mathbf{D}_t \geq q_a\} \, \mathrm{d}t \right)^{-1} \int_0^T \mathbf{D}_t \, \mathbb{1}\{\mathbf{D}_t \geq q_a\} \, \mathrm{d}t \tag{1.1}$$

where q_α is the α-quantile of the $\mathbf{D}_t$ and $\mathbb{1}$ is the indicator function which is 1 if the statement in brackets is true and 0 otherwise.[3]

In the discrete case, the CDD is defined as

$$\text{CDD}_\alpha(S) = \frac{1}{|D^\alpha|} \sum_{i \in D^\alpha} D_i, \qquad D^\alpha = \{j \mid D_j \geq q_\alpha, j \in \{1, \ldots, N\}\}, \tag{1.2}$$

where D^α includes all drawdowns which are large enough to be among the worst $(1 - \alpha) \cdot 100\%$, q_α is the α-quantile of the D_i and $|\cdot|$ denotes set cardinality.

Therefore, the MDD corresponds to $\alpha \nearrow 1$ and the ADD to $\alpha \searrow 0$. The lower the α, the more of the drawdown curve is taken into account.

A conceptually similar drawdown measure is the Conditional Expected Drawdown (CED) advocated by Goldberg and Mahmoud (2017). The CED is defined as the tail mean of the

---

[3]We omit some intricate notation from Chekhlov et al. (2005) by assuming that the reader is familiar with the concept of quantiles. Moreover, in the simulations below, the correction term in the CDD formula in Chekhlov et al. (2005) will diminish due to large sample sizes.

Maximum Drawdown distribution. If $q_\alpha^{\mathcal{M}(\mathbf{S})}$ is the α-quantile of the MDD distribution of the price process $\mathbf{S}$, then

$$\text{CED}_\alpha(\mathbf{S}) = \mathbb{E}\left[\text{MDD}(\mathbf{S}) \,|\, \text{MDD}(\mathbf{S}) \geq q_\alpha^{\mathcal{M}(\mathbf{S})}\right]. \tag{1.3}$$

To obtain formulas for the discretized process, we can simply substitute S for $\mathbf{S}$ in the above formula.

To compute the realized CDD for a single sample path, the theoretical quantile q^α is replaced by the empirical quantile of the D_i. However, the realized CED for a single sample path cannot be computed because it does not exist. From a single realized Maximum Drawdown inference of the Maximum Drawdown distribution's quantiles and conditional expectation is not feasible.

When we are concerned with forward-looking risk assessment for the discretized process, we usually make distributional assumptions such that we can sample repeatedly from the stochastic price process S. Given a set of J samples $\mathcal{S} = \{S^j\}_{j=1,\ldots,J} = \{(S_1^j, \ldots, S_N^j)\}_{j=1,\ldots,J}$ from the discretized process, we can estimate the expected MDD straightforwardly by the sample average of the MDDs $\frac{1}{J}\sum_{j=1}^J \text{MDD}(S^j)$. We estimate the expected ADD and expected CDD analogously. For the expected CED, we estimate

$$\frac{1}{|\text{MDD}^\alpha|} \sum_{j \in \text{MDD}^\alpha} \text{MDD}(S^j), \tag{1.4}$$

where $\text{MDD}^\alpha = \left\{j \,\middle|\, \text{MDD}(S^j) > q_\alpha^{\mathcal{M}(\mathcal{S})}, j \in \{1, \ldots, J\}\right\}$ and $q_\alpha^{\mathcal{M}(\mathcal{S})}$ is the sample quantile of the MDD distribution.

1.2.3 A new end-of-period drawdown measure

In practice, investors often enter into investments with a certain investment horizon in mind. Similarly, portfolio managers' performance is usually evaluated at fixed dates and trading strategies are implemented to be rebalanced after regular holding periods. Therefore, investors, portfolio managers and traders are interested in choosing their positions such that the performance at the end of the relevant period is as desired.

However, as argued before, the path between now and the invest horizon T may still be relevant for two reasons. First, liquidity and other constraints may require the path to remain above a certain threshold. Second, the path may significantly influence how performance is evaluated psychologically at time T. For example, consider the two strategies in figure 1.1 which are constructed as realizations from the same stochastic process and have the same final values.
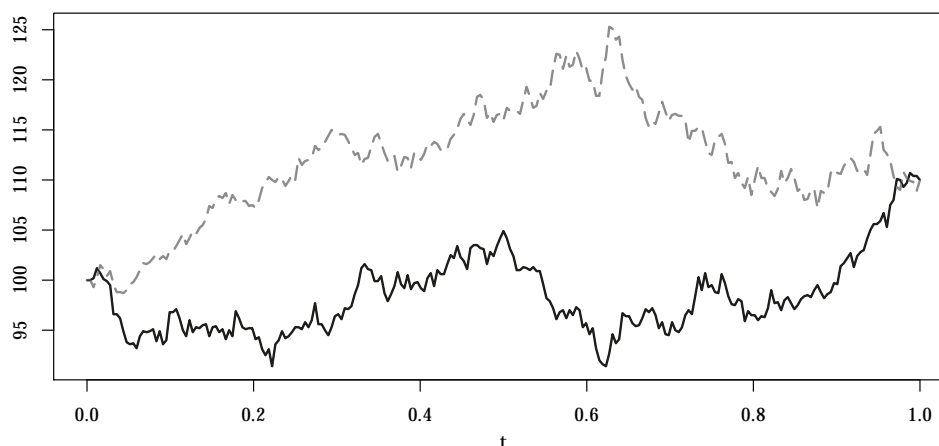
Figure 1.1: Sample paths of two strategies stemming from the same stochastic process with the same Maximum Drawdown of 15% and the same final value.

At time T, an investor in the grey strategy (dashed line) is likely to be disappointed with the strategy because of a significant loss in value compared to the previously achieved high. On the contrary, an investor will probably be fairly satisfied with the black strategy since it has earned significantly from a previous slump.

Although the risk measures CDD and CED take the path into account, they cannot distinguish between the grey and the black strategy as, for example, the Maximum Drawdown for both strategies is exactly the same in figure 1.1.[4] This is not to say that CDD or CED disregard time; in fact, the expected Maximum Drawdown scales with time because the likelihood of an extreme drawdown event increases. Choosing a strategy with low expected CDD aims to control the likelihood of (prolonged) significant drawdown periods. Choosing strategies with low expected CED tends to ensure that the expected magnitude of an extreme Maximum Drawdown event is small.

Nonetheless, for all the above-mentioned drawdown risk measures, a large loss within the sample path always facilitates a large drawdown irrespective of a subsequent rally. For example, consider a strategy for which short-term outlooks are dire but long-term forecasts are excellent. The Maximum Drawdown – and all of the drawdown measures defined so far – will deem both the long-term and the short-term strategy very risky due to the initial slump. However, it is obvious that in this case the time horizon of the investor is essential to determine if the strategy is risky

---

[4]As argued in section 1.2, the CED cannot be computed for a single sample path. Nonetheless, for processes with sample-wise indistinguishable Maximum Drawdowns, the CEDs must also be indistinguishable because the CED is a function of the MDD distribution.

or not. Therefore, we want to define a measure that adequately takes the time horizon into account.

This measure, we call it end-of-period drawdown, should capture well-esta blished behavioral aspects. Like all drawdown measures, it should focus on losses because of the well-known insight from Prospect Theory that losses are weighed significantly higher than gains, see Kahneman and Tversky (1979), Kahneman et al. (1990) and Veld and Veld-Merkoulova (2008). The reference point with respect to which losses shall be measured is the running maximum of the path – a natural choice since the maximum could have been achieved by optimal market timing, i.e., switching from the risky strategy to holding cash at the strategy's maximum. Therefore, any outcome below the running maximum can be perceived as a loss. Since the running maximum is time-variant, losses are measured with respect to a dynamic benchmark.[5] The severity of the loss is measured by the loss incurred from the peak to the final position relative to the initial investment, which allows straightforward comparisons of strategies with different initial investments.

We define the end-of-period drawdown as

$$\text{eopDD}(\mathbf{S}) = \frac{\mathbf{M}_T - \mathbf{S}_T}{\mathbf{S}_0}, \tag{1.5}$$

where $M_T = \sup_{u \in [0,T]} S_u$ and T is the investment horizon as before. For the discretized process, we define $\text{eopDD}(S) = \frac{M_N - S_N}{S_0}$. The drawdown measures can be interpreted intuitively: High drawdown indicates high risk, i.e., a large loss in portfolio value from the maximum to the end.

The drawdown thus defined has a straightforward interpretation in terms of regret with respect to market timing. In the regret literature[6], regret is defined as the difference between two strategies: the ex-post optimal strategy and the strategy that was actually pursued (Gollier, 2020). The larger the difference between the two strategies, the larger the disutility for the investor. Consider an investor who can shift his money between a risky strategy and risk-free cash holdings. An

---

[5]For an application of reference point adaption in the portfolio context as well as more literature on reference point dynamics refer to Shi et al. (2015).

[6]Early accounts of regret in decision-making include Bell (1982) and Bell (1983). Work on behavioral finance frequently emphasizes the role of regret in the finance context, for example, Shefrin and Statman (1984), Shefrin and Statman (1985), Barberis et al. (2001), Lin et al. (2006), and Strahilevitz et al. (2011). More recent research efforts involving regret include Bleichrodt et al. (2010), Bleichrodt and Wakker (2015), Gollier (2020), and Diecidue and Somasundaram (2017). Neural evidence for regret is presented by Frydman and Camerer (2016) in an experimental asset market.

investor with optimal market timing would invest into the risky strategy and exit into cash at the risky strategy's maximum. The end-of-period drawdown captures the regret of an investor who cannot time the market perfectly but compares his risky strategy to the optimal strategy at the end of the time horizon. By elementary calculation, we can express the end-of-period drawdown as

$$\text{eopDD}(\mathbf{S}) = \underbrace{\frac{\mathbf{M}_T - \mathbf{S}_0}{\mathbf{S}_0}}_{r_1} - \underbrace{\frac{\mathbf{S}_T - \mathbf{S}_0}{\mathbf{S}_0}}_{r_2}, \tag{1.6}$$

i.e., as the difference of the discrete relative return of the ex-post best strategy r_1 with optimal market timing and the return of the chosen strategy r_2. Therefore, the eopDD captures the return-wise difference between the two strategies. Even if the investor is not aware of the running maximum as a reference point, his evaluation of investment success may still be influenced by it. In an experimental study, Unser (2000) finds that subjects are not necessarily aware of their reference points.

To dissect the influences on the expected end-of-period drawdown in more detail, we want to discriminate whether an increase in $\mathbb{E}[\text{eopDD}]$ is driven by an increased likelihood of a drawdown event or an increased severity given a drawdown event. To this end, we consider the following decomposition which is immediately justified by the definition of the conditional expectation:

$$\mathbb{E}\left[\text{eopDD}\right] = \mathbb{E}\left[\text{eopDD} \,|\, \text{eopDD} > 0\right] \cdot \mathbb{P}\left(\text{eopDD} > 0\right). \tag{1.7}$$

This decomposition allows us to attribute changes in eopDD either to the severity or likelihood of a drawdown. The concept of focusing on the probability of drawdown can be found in Angoshtari et al. (2016).

Despite evident similarities, the eopDD can be substantially different from already existing drawdown measures. In fact, the difference between the eopDD and the Maximum Drawdown can be arbitrarily large, for example, if a large drop in portfolio value is entirely offset by a subsequent rally, such that MDD is large and eopDD is zero.

Several sets of properties (e.g., coherence (Artzner et al., 1999)) have been proposed in the literature to ensure that risk measures are sensible. For drawdown-based measures, the class of (generalized) deviation measures originally proposed by Rockafellar et al. (2006) and generalized to path-dependent measures by Goldberg and Mahmoud (2017) is most suitable. In general terms, for price processes $\mathbf{S}$ from a space of stochastic processes $\mathcal{R}^\infty$ a risk measure $\rho \colon \mathcal{R}^\infty \to \mathbb{R}$ is a generalized deviation measure if for all $\mathbf{S}, \tilde{\mathbf{S}} \in \mathcal{R}^\infty$ and all constant deterministic $\mathbf{C} \in \mathcal{R}^\infty$:

(C1) Normalization: $\rho(\mathbf{C}) = 0$

(C2) Non-negativity: $\rho(\mathbf{S}) \geq 0$

(C3) Shift invariance: $\rho(\mathbf{S} + \mathbf{C}) = \rho(\mathbf{S})$

(C4) Positive degree-one homogeneity: $\rho(\lambda \mathbf{S}) = \lambda \rho(\mathbf{S})$ for all $\lambda > 0$

(C5) Convexity: $\rho(\lambda \mathbf{S} + (1 - \lambda)\tilde{\mathbf{S}}) \leq \lambda \rho(\mathbf{S}) + (1 - \lambda)\rho(\tilde{\mathbf{S}})$ for all $\lambda \in [0, 1]$.

The eopDD is a generalized deviation measure since the conditions (C1)–(C4) hold trivially and for condition (C5) it suffices to see that $\max_{t \in [0,T]} f(t) + g(t) \leq \max_{t \in [0,T]} f(t) + \max_{t \in [0,T]} g(t)$ which holds for general functions f and g with $f(t) + g(t) < \infty$ for all $t \in [0, T]$.[7]

Risk measures from the CDD family and from the CED family are also generalized deviation measures according to Chekhlov et al. (2005) and Goldberg and Mahmoud (2017), respectively. Thus, all drawdown measures under consideration satisfy these desirable properties.

1.2.4 Comparing historical performance – a case study

Theoretical properties aside, we want to see how the end-of-period drawdown performs in practice. To this end, we conduct a case study on portfolio data by considering the following setting: Let there be 100 portfolio managers who manage portfolios assembled from the S&P 500 stock index from January 1990 until December 2017. Each manager picks 100 stocks at random and assigns equal weights. All portfolios have a turnover of 20% each year and are adjusted for stocks entering and exiting the S&P 500 index.[8] For each year and each portfolio manager, we compute the MDD, $\text{CDD}_{0.8}$, ADD and eopDD.

In figure 1.2, boxplots of the drawdown quantities are displayed to give a rough idea about the scale of the different drawdown measures. For each drawdown measure, the results for all managers and all years are combined. Not surprisingly, medians, upper and lower quartiles and whiskers decrease from MDD over $\text{CDD}_{0.8}$ to ADD. Compared to ADD, the eopDD has an even lower median but higher variance. While this is unlikely for the other drawdown measures, some of the eopDDs are exactly zero when the stock price at year-end constitutes an annual maximum.

---

[7]As in Goldberg and Mahmoud (2017), the conditions are applied to the risk measure defined in absolute terms, i.e., $\rho(\mathbf{S}) = \mathbf{M}_T - \mathbf{S}_T$. For the eopDD defined in relative terms, the conditions (C3) and (C4) are replaced by a new multiplicative scalability condition $\rho(\lambda \mathbf{S}) = \rho(\mathbf{S})$ for $\lambda > 0$.

[8]In detail, we first drop the stocks which left the index, then randomly eliminate stocks to arrive at 20% and then draw new stocks from the current index constituents which have not been eliminated in the previous step.
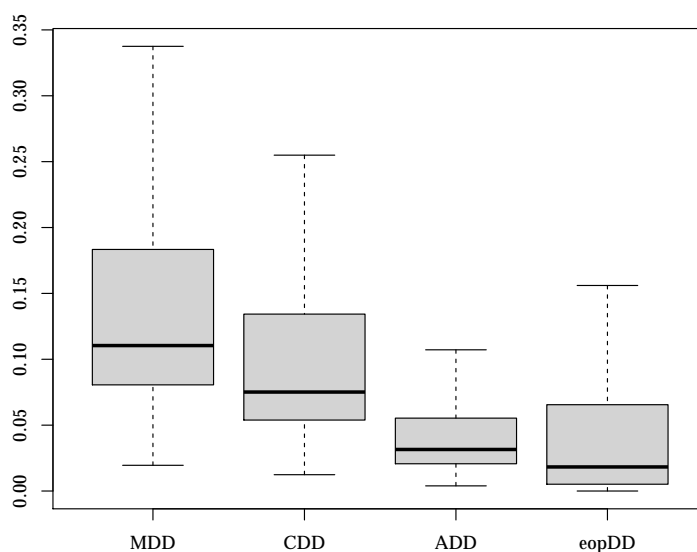
Figure 1.2: Boxplots as summary statistics for each drawdown measure, each year and 1,000 portfolio managers. From left to right, the drawdown measures MDD, $CDD_{0.8}$, ADD and eopDD are displayed. Outliers are omitted.

To analyze whether the different drawdown measures provide distinct information, we compute rank correlations. I.e., we rank the 100 portfolios by each of the four drawdown measures and compute the degree of similarity between the rankings. Once annual rank correlations are obtained for each pair of drawdown measures, we average the annual results and present the rank correlations for each pair of drawdown measures in table 1.1. If two drawdown measures rank the 100 portfolios in equal order for every manager and every year, the rank correlation should be 1. If two drawdown measures yield strictly contrary rankings, the rank correlation would equal -1.

Most notably, we observe substantial differences between the end-of-period drawdown and all other drawdown measures in table 1.1. The rank correlations between eopDD and one of MDD, $CDD_{0.8}$, and ADD are approximately 0.3 – the rank correlations between MDD, $CDD_{0.8}$, and ADD are each at least twice as large. Hence, we infer that the eopDD measures something distinct and does not simply reproduce the results from previous drawdown measures.

Next, recall that MDD and ADD are members of the CDD family where MDD corresponds to α approaching 1, ADD corresponds to α approaching 0. Consistently, the rank correlations between either MDD or ADD with $CDD_{0.8}$ are higher (0.80 and 0.76, respectively) than the rank correlation between both extremes, i.e., between MDD and ADD (0.64). In general, we obtain rank correlations well above zero for all drawdown measures.

Methodologically, we proceed as in Eling and Schuhmacher (2007) who compare drawdown-based

	eopDD	MDD	$\text{CDD}_{0.8}$	ADD
eopDD	1	—	—	—
MDD	0.25	1	—	—
$\text{CDD}_{0.8}$	0.30	0.76	1	—
ADD	0.30	0.64	0.80	1

Table 1.1: Rank correlations (average annual Kendall's τ-b) between portfolios ranked on different drawdown measures. Upper triangle is omitted because of symmetry.

performance measures using Spearman's rank correlations. Instead of Spearman's ρ, we compute the rank correlations using Kendall's τ because it corrects more accurately for ties, i.e., when two (or more) portfolios are assigned the same risk measure (Kendall, 1945). Specifically, we apply Kendall's τ-b as in Agresti (2010). Correcting for ties is advisable in this context because the eopDD is more likely to be exactly zero than the other drawdown measures, thus inducing ties. If we repeat the analysis and use Spearman's ρ instead, all rank correlations are slightly higher and all qualitative results remain unaltered. Moreover, the results are robust to reasonable changes to the setting, i.e., more portfolio managers (e.g., 1,000), different portfolio sizes (e.g., 50) and different turnover (e.g., 10% or 30%).

Hence, we find that while the eopDD and the other drawdown measures are related – as indicated by the positive rank correlations of about 0.30 – the eopDD does not coincide with the other drawdown measures which display much higher rank correlations among themselves.

1.3 Simulation Framework

We want to analyze how return moments influence drawdown risk measures. Because drawdown measures are path-dependent, we need a simulation model which generates entire sample paths. Once such sets of paths $\mathcal{S}$ are sampled, we compute the drawdown risk measures with the formulas from section 1.2. We provide details on the simulation setup here and the simulation results in section 1.4.

The most important ingredient for the model is a probability distribution whose moments can be manipulated in a flexible way. The normal inverse Gaussian distribution (NIG) is an intuitive choice in this regard because its expectation, variance, skewness, and kurtosis – i.e., its first four moments – can be varied extensively. The NIG distribution belongs to the class of

generalized hyperbolic distributions and was introduced into the mathematical finance literature by Barndorff-Nielsen (1997). Its density is given by

$$f(x, \alpha, \beta, \mu, \delta) = \frac{\alpha \delta \exp\left(\delta\gamma + \beta(x - \mu)\right)}{\pi\sqrt{\delta^2 + (x - \mu)^2}} K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right), \qquad (1.8)$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $K_1(u) = \frac{1}{2}\int_0^\infty \exp\left(-\frac{1}{2}u\left(v + \frac{1}{v}\right)\right) \mathrm{d}v$ is the modified Bessel function of the third order and index 1 (Kalemanova et al., 2007). The density is defined for $0 \leq |\beta| < \alpha$, $\delta > 0$ and $\mu \in \mathbb{R}$. Intuitively, α governs the tail heaviness and β governs asymmetry, μ is the location parameter and δ the scale parameter.

The normal inverse Gaussian distribution has been applied in several areas of finance, for example, in Eriksson et al. (2009), Kalemanova et al. (2007), and Homm and Pigorsch (2012), mainly to model returns in financial time series, see Aas et al. (2005), Barndorff-Nielsen and Stelzer (2005) and Wilhelmsson (2009) and the references therein. Log returns can be fitted especially well by the NIG distribution (Barndorff-Nielsen, 1998). Consequently, we use the NIG distribution to model the log returns. Whenever we mention returns henceforth, we mean log returns $\log(S_{t+1}/S_t)$.

To proceed from a distribution to a process, we exploit that the normal inverse Gaussian family is closed under convolution: if $Y_1, Y_2 \sim \mathrm{NIG}(\alpha, \beta, \mu_{1,2}, \delta_{1,2})$ are independent, then $Y_1 + Y_2 \sim \mathrm{NIG}(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2)$ (Barndorff-Nielsen et al., 2013). By infinite divisibility, we obtain the normal inverse Gaussian Lévy process as in Barndorff-Nielsen (1997) which has stationary and independent increments by definition and has normal inverse Gaussian marginals at all times.[9]

Once the returns $r_t \sim \mathrm{NIG}(\alpha, \beta, t\mu, t\delta)$ are simulated, the price process can be computed by $S_t = S_0 \exp(r_t)$ where S_0 is the price at time 0.[10] By construction, the annual return of this price process is distributed $\mathrm{NIG}(\alpha, \beta, \mu, \delta)$.

Therefore, we can generate sample paths of the NIG Lévy process with the desired moment properties if we can find a NIG random variable with the same moments. In other words, for a sample path to, on average, have annual returns with some specified expectation, variance, skewness, and kurtosis, it suffices to find parameters α, β, μ and δ such that a NIG random variable with these parameters has the same four moments. The moment formulas[11] for $X \sim$

---

[9]In detail, the marginal distribution of the process is $\mathrm{NIG}(\alpha, \beta, t\mu, t\delta)$ at time t (Barndorff-Nielsen and Stelzer, 2005).

[10]The price process hence belongs to the class of exponential Lévy processes.

[11]The standard deviation is the square root of the centralized second moment. Skewness and kurtosis are also centralized and normalized. All definitions are provided in A.1.

NIG($\alpha, \beta, \mu, \delta$) are given by Kalemanova et al. (2007):

$$
\begin{aligned}
\mathbb{E}[X] &= \mu + \frac{\delta\beta}{\gamma} \\
\mathrm{sd}(X) &= \sqrt{\frac{\delta\alpha^2}{\gamma^3}} \\
\mathrm{skewness}(X) &= \frac{3\beta}{\alpha\sqrt{\delta\gamma}} \\
\mathrm{kurtosis}(X) &= 3 + 3\left(1 + 4\frac{\beta^2}{\alpha^2}\right)\frac{1}{\delta\gamma}.
\end{aligned}
\tag{1.9}
$$

In this non-linear system of four equations and four unknowns, we fix left sides and solve for the free parameters α, β, μ, and δ using Broyden's Quasi-Newton method and the double dogleg trust region method (Dennis and Schnabel, 1996). For example, if an expected return of 8%, a standard deviation of 20%, a skewness of 0 and a kurtosis of 4 are the choices of the return moments, then the parameters which generate such a process are $\alpha \approx 8.66$, $\beta = 0$, $\mu = 0.08$, and $\delta \approx 0.35$.

Implementing the NIG Lévy process for discretized paths is straightforward. Consider the one-year case with 252 trading days:

1. Fix the desired four moments of annual returns.

2. Solve the system of equations (1.9) to obtain the parameters α, β, μ, and δ.

3. Draw 252 realizations from the NIG($\alpha, \beta, \mu/252, \delta/252$) distribution and obtain the return process by their cumulative sum.

4. Take the exponential and multiply by S_0 to obtain the price process.

5. Use the price path to compute the different drawdown measures from section 1.2.

To be suitable for such an analysis of moment sensitivity, a stochastic model for the asset returns needs to exhibit specific properties. First, it must be flexible enough to allow variation in each of the moments while the others are held constant. Otherwise, computing meaningful comparative statics is infeasible. For example, in a standard geometric Brownian motion without jumps it is possible to vary the two parameters (μ and σ), but skewness and kurtosis cannot be modified separately. Second, paths generated from the model should resemble actual paths in financial data. This aspect is intertwined with the first, as stylized facts about asset returns like left-skewness and fat tails immediately relate to moment properties. Third, since the moments are fixed for the annual return but simulated in daily steps, it is of great advantage if distributions are

closed under convolution. Similarly, analytical solutions for the first four moments should exist such that systems of equations like the above can be formulated. Theoretically, the third and fourth requirements could be relaxed if parameters were found by simulation. However, immense computing power would be required to reduce computational imprecision to decent levels. Last, catering mostly to elegance, we would like the solutions of the system of equations like (1.9) to be unique, i.e., for a given set of moments there is at most one process in this family that generates a process with these return moments.

From the myriad of stochastic processes, the NIG Lévy process satisfies all of the above criteria: The first four moments can be steered independently in a flexible manner, the process has been found to resemble financial return data (see sources above), analytical solutions for the first four moments obviously exist, the NIG family is closed under convolution as noted above and a proof of uniqueness is provided in A.4. We cannot rule out the existence of another equally suited model but we considered a wide range of models – including, for example, all models in Schuhmacher and Eling (2011) – and found none meeting all of the criteria.[12]

In order to still test robustness of the NIG results, we have devised a tailor-made jump diffusion model which meets the above criteria. It is, however, less flexible with respect to the moments attainable. Its design is detailed in section 1.5.

In the NIG model, not all combinations of moments are attainable, i.e., the system of equations need not have a solution for all combinations of prescribed moments. This is not surprising because, for example, high kurtosis and low standard deviation do not coexist well. We find virtually no restrictions to choose economically sensible expectations and standard deviations. For the skewness, symmetrical intervals around zero are attainable due to the parameter restriction $|\beta| < \alpha$. For the kurtosis, arbitrary values are accessible as long as they are strictly greater than 3.[13] Since the economically interesting case of large kurtosis corresponds to values larger than 3, the latter condition is, practically, no constraint.

---

[12]Many distributions can be disregarded right away because of too few parameters, for example, the normal, gamma, logistic, Weibull, student t, uniform distributions etc. In more detail, we considered, for example, Lévy processes with beta, extended skew-normal and extended skew-t distributed increments, the variance gamma process as well as other generalized hyperbolic distributions and the jump diffusion processes of Merton with normally distributed jumps and the jump diffusion model by Kou (2002) with double-exponential jumps.

[13]This is clear from the kurtosis formula in (1.9) because $\delta, \gamma > 0$. Generally, all models which can be generated as time-changed Brownian motion with drift (like the NIG model, cf. Barndorff-Nielsen (1997)) are leptokurtic if the subordinator is not deterministic (Cont and Tankov, 2004).

To validate that the simulation is well-behaved, we simulate 1,000,000 sample paths and compute the moments of the realized returns for each path. If the simulation framework is correct, the difference between the prescribed moments and the average realized moments should approach zero – which it does. Thus, the model, on average, generates paths with the prescribed return moments.

1.4 Drawdown Sensitivities to Return Moments

To analyze the effects of return moments on drawdown measures, we conduct a simulation study based on the model from section 1.3. In this model, it is possible to specify the first four moments and generate sample paths which in expectation have these return moments. To dissect the effects of the different moments on the drawdown, we choose a base case scenario and then vary each of the moments one at a time, i.e., we simulate comparative statics. While either expectation, standard deviation, skewness, or kurtosis of returns are varied, we compute the expected drawdown measures from section 1.2, i.e., expected Conditional Drawdown and Conditional Expected Drawdown, expected end-of-period drawdown, and its decomposition.

To be precise, there are slight differences in estimation between the risk measures due to their definitions. For the CDD family, the eopDD and the conditional eopDD, estimators of the expectations of the drawdown measures are computed. Canonically, the expectations are estimated via the sample mean. For the CED family and the probability of end-of-period drawdown, one directly computes an estimator for the drawdown measure. Canonically, the quantile is estimated by the sample quantile and the probability is estimated by the sample frequency.

For each comparative static and for each parameter combination, we employ the model to generate 1,000,000 paths from which, in turn, all risk measures are computed. The investment horizon T is fixed to one year following Benartzi and Thaler (1995) who identify one year as an investor's typical evaluation period. To mimic trading-daily observations, the path is discretized into 252 equidistant time intervals. The strategy is arbitrarily initialized at $S_0 = 100$ and we consider annual logarithmic returns. As a base case, we consider an expected return of 8% p. a., a standard deviation of 20% p. a., a skewness of 0 and a kurtosis of 4. The expected return and standard deviation resemble rough estimates of stock market characteristics and are close to the choices in Benartzi and Thaler (1995). The choices of skewness and kurtosis induce neither right

nor left skewness and avoid much excess kurtosis in the base case such that these effects can be introduced separately.[14]

From the CDD family, we investigate the MDD and the ADD – which are the two extreme cases – and the $CDD_{0.8}$ as an example in between. From the class of CED measures, we consider the CED with confidence levels $0.95, 0.90, 0.85$, and 0.8. Intuitively, for $\alpha = 1$ only the single largest Maximum Drawdown would be considered, making the measure possibly vulnerable to outliers. Since the rationale behind the CED is to control the maximal loss for most possible scenarios, rather high confidence levels make sense economically.

1.4.1 Variation of return moments

We present the main results of the simulation study in figures 1.3 and 1.4. All plots can be read intuitively in a similar fashion. The first column of figure 1.3 contains the comparative static for the first moment. Here, the other moments – standard deviation, skewness, and kurtosis – remain fixed at the base case level while the expected return on the x-axis is varied between 0 and 20%. The three plots then display the expected drawdown measures from the different drawdown families. According to figure 1.3, all drawdown measures decline monotonically with rising expected return. The effect is quite significant with some drawdown measures declining by almost half. These results are plausible since strategies with positive trend tend to make up intermediate drawdowns such that the drawdown graph is pushed to zero frequently. Hence, other moments being equal, the higher the expected return, the lower the drawdown measures.

The comparative static for the second moment is displayed in the right column of figure 1.3. We find that all drawdown measures are monotonically increasing in the standard deviation. The size of the increase is large – both in absolute terms where all measures increase almost threefold over the interval between 10% and 30% and in relative terms compared to the other moment effects.[15] As for the first moment, the general tendency of the effect is intuitive: A higher standard deviation induces more and larger fluctuations in the paths which, in turn, facilitate more frequent and large drawdown events. Hence, all else equal, higher standard deviation translates into higher drawdown measures.

---

[14]An excess kurtosis of 0 is not feasible in the model because the kurtosis has to be strictly greater than 3. Additionally, for a kurtosis very close to 3, the range of admissible values for the skewness is very small because the range increases with higher kurtosis. Hence, we deem 4 to be a sensible base case.

[15]The y-axes in figures 1.3 and 1.4 are the same for each drawdown family such that results for the different moments can easily be compared in size.
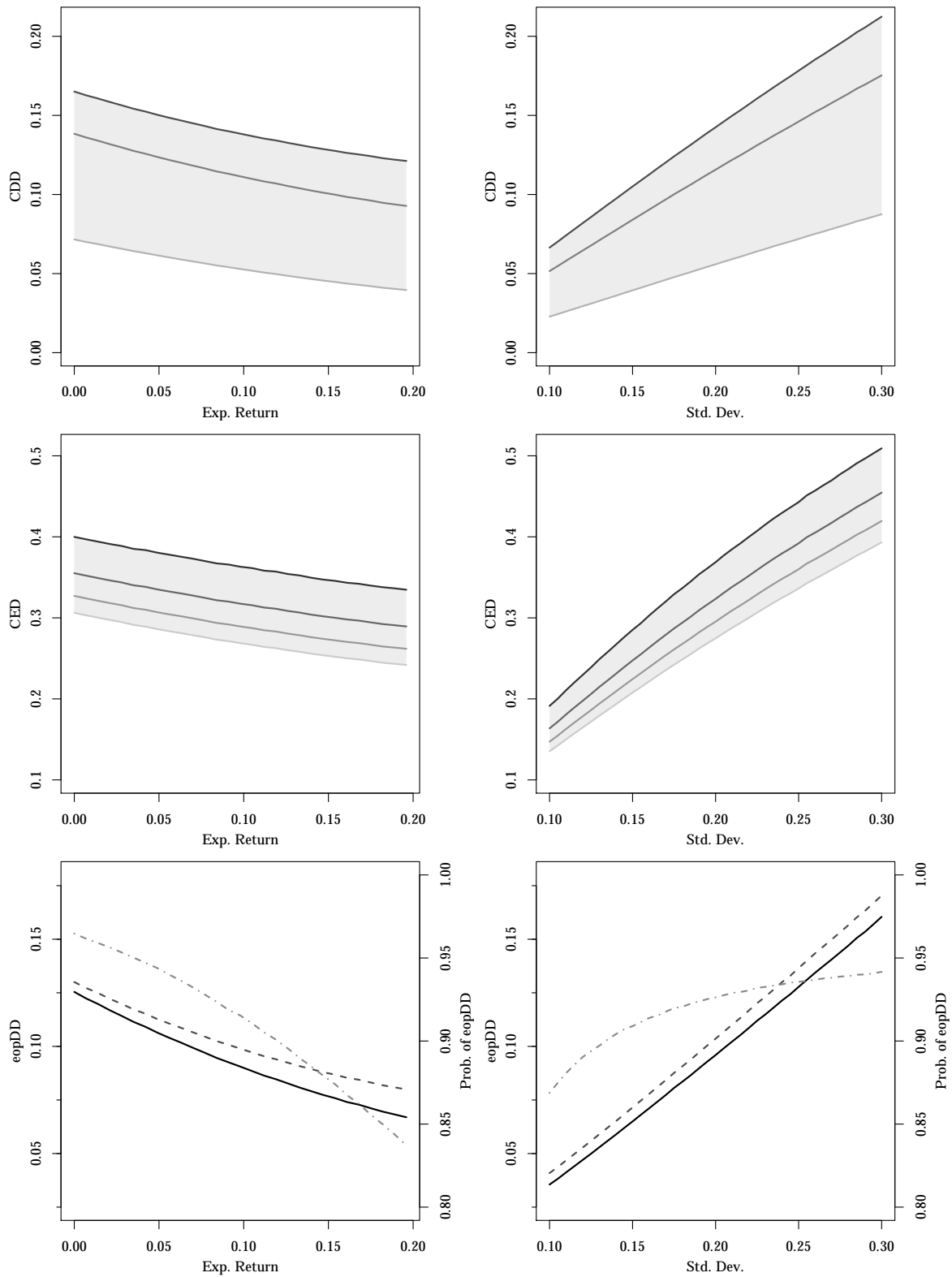
Figure 1.3: Comparative static for 1st moment (left column) and 2nd moment. The CDD-charts (first row) contain the MDD, $CDD_{0.8}$ and ADD from top to bottom. In the middle row, we display the CED for levels $0.95, 0.9, 0.85$, and 0.8. The bottom row contains the eopDD (bold line) and its decomposition into the probability of eopDD (dot-dashed line) and the conditional eopDD (dashed line). All paths are simulated from the normal inverse Gaussian Lévy model.

When it comes to the skewness effects, which are depicted in the left column of figure 1.4, we observe the general trend that left-skewness induces higher drawdown and right-skewness induces lower drawdown. Since left-skewness indicates the presence of more extreme losses, higher drawdown measures seem intuitive. The same is true for the right-skewness by the analogous argument. However, there are numerous differences between the drawdown families: For the Maximum Drawdown and $CDD_{0.8}$, the relationship is monotonic but almost flat for positive values of skewness. For the Average Drawdown, the relationship is almost flat and monotonicity is lost. On the contrary, the effect of skewness on the CED is strong across the board with the effect of varying the skewness between $[-0.75, 0.75]$ being significantly larger than the effect of varying the expectation between $[0, 0.20]$. Last, the expected end-of-period drawdown and its conditional version behave like members of the CDD family. While the eopDD is as flat and non-monotonic as the Average Drawdown, the conditional eopDD slightly follows the trend like the Maximum Drawdown. In summary, we observe a general trend for higher drawdown with increasing left-skewness and vice versa, which however affects different drawdown measures to different degrees.

The right column of figure 1.4 contains the results on the kurtosis effects. If the kurtosis increases, i.e., the distribution becomes more heavy-tailed, the CDD decreases monotonically and quite significantly in size. On the contrary, changes in kurtosis leave the CED almost unaffected. For high confidence levels, the CED increases slightly; for low confidence levels, it decreases slightly; and it is mostly flat in between. Compared to the other moments, the effect of the kurtosis on the CED is almost negligible. The eopDD indicates less risk for higher kurtosis just like the CDD. The same holds for the elements of its decomposition which also decrease monotonically in kurtosis. Hence, we observe a tendency for CDD and eopDD to decrease when kurtosis increases and for CED to remain unaffected.

1.4.2 Drawdown measures and moment preferences

All else equal, what would you choose: An investment with high returns or an investment with low returns? An investment with a chance of positive extremes or an investment with a chance of negative extremes? Intuitively, choosing the investment with high expectation and skewness, i.e., high odd moments, seems like a good idea. Again, all else equal, what would you choose: An investment with high variability or low variability? With more likely extreme events or less likely extreme events? In this case, intuition is likely to point to the choice with low standard deviation and kurtosis, i.e., low even moments. In general, there is both theoretical and empirical evidence
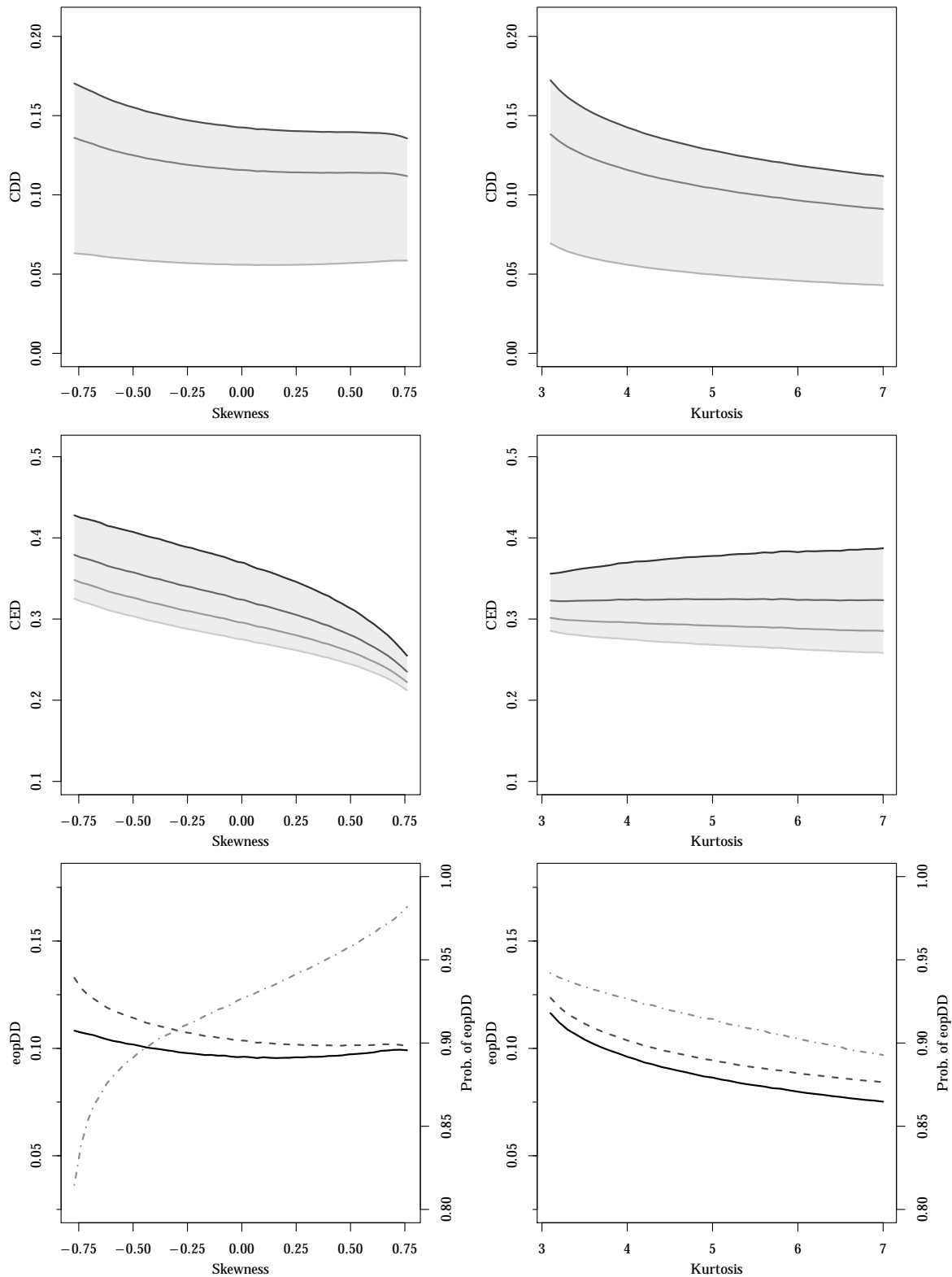
Figure 1.4: Comparative static for 3rd moment (left column) and 4th moment. The CDD-charts (first row) contain the MDD, $CDD_{0.8}$ and ADD from top to bottom. In the middle row, we display the CED for levels $0.95, 0.9, 0.85$, and 0.8. The bottom row contains the eopDD (bold line) and its decomposition into the probability of eopDD (dot-dashed line) and the conditional eopDD (dashed line). All paths are simulated from the normal inverse Gaussian Lévy model.

that investors usually prefer assets with high odd moments and low even moments, see Scott and Horvath (1980), Harvey and Siddique (2000), and Dittmar (2002).

Do drawdown measures reflect this pattern? We address this question using the comparative statics in figures 1.3 and 1.4. For example, all drawdown measures are consistent with the moment preferences for the first moment according to figure 1.3 because if the first moment increases, all drawdown measures decrease monotonically. Since quantitative statements are not warranted theoretically, we confine ourselves to observing the drawdown measures' general tendencies.

If drawdown measures do not act accordingly, there may be unintended consequences. For example, aiming for a strategy with low drawdown risk, investors may end up taking on high skewness or kurtosis risk. By looking at the directions in the plots, we can investigate if drawdown measures react favorably to high odd moments and dismissively to high even moments, vice versa, or not at all. To summarize the results from the comparative statics, table 1.2 lists the different drawdown measures' directions.

	$\mathbb{E}[\text{CDD}]$	CED	$\mathbb{E}[\text{eopDD}]$
Expectation up	down	down	down
Standard Deviation up	up	up	up
Skewness up	down/constant *	down	constant
Kurtosis up	down	approx. constant **	down

Table 1.2: Summary of the trends in the comparative statistics of figures 1.3 and 1.4. Risk measures consistent with standard moment preferences would yield columns of the form (down, up, down, up). The first column contains the return moments.

* Down for high confidence levels, approx. constant for lower confidence levels.

** Slightly up for high confidence levels, slightly down for low confidence levels.

For the first two moments, all drawdown measures agree with the standard moment preferences unanimously. The CED perfectly matches the third moment preference because when skewness increases it declines strongly and monotonically. The effects for the kurtosis are almost negligible in comparison. However, for high α, i.e., when only the largest Maximum Drawdowns are taken into account, the CED increases monotonically with the kurtosis. Thus, it is consistent with all four moment preferences. The CED for low α declines slightly with kurtosis.

For the expected CDD, the results for the skewness differ according to the confidence level. While

the MDD mostly agrees with a preference for higher skewness, the ADD is almost constant. Turning to the kurtosis, all CDD measures decline significantly, which implies that heavier tails indicate lower risk – something that is at odds with the moment preferences. The results for the expected eopDD very closely resemble those of the ADD from the CDD family.

In conclusion, using the CED with high confidence levels appears to capture the standard moment preferences best. The MDD (along with other members of the CDD family with high α) is at least in accordance with preferences for the first three moments.

1.5 Robustness

We want to analyze how return moments like skewness and kurtosis influence drawdown risk measures. The model we have employed – the normal inverse Gaussian Lévy process – is merely a tool to analyze this relationship. Naturally, the question arises to what extent the results depend on the model. To address this potential concern, we repeat the analysis of section 1.4 with an alternative model which is quite different in nature, namely a jump diffusion model similar to Merton's model (Merton, 1976) but with a discrete jump height distribution.

In general, jump diffusion models generate the price process $\mathbf{S} = \{S_t\}_{t \in [0,T]}$ via

$$S_t = S_0 \exp\left(\mu t + \sigma W_t + \sum_{i=1}^{N_t} Y_i\right), \tag{1.10}$$

where μ and $\sigma > 0$ are the drift and volatility of the diffusion component, respectively, W_t is a standard Brownian motion, N_t is a Poisson distributed random variable with parameter λt, and the Y_i are independent and identically distributed (Cont and Tankov, 2004). All random components are pairwise independent and the sum in the exponent is sometimes referred to as a compound Poisson process. These models have both been extensively employed in financial modeling and studied theoretically, cf. Ball and Torous (1983), Kou (2002), Cont and Tankov (2004), and Hanson (2007). The economic reasoning behind the model is this: small continuous fluctuations are modeled by a geometric Brownian motion and the occurrences of abnormal fluctuations due to significant new information are modeled by a Poisson jump process (Ball and Torous, 1983). This way, the model can represent typical features of real stock price data like asymmetry and excess kurtosis while still being computationally manageable and economically sensible. It belongs to the class of exponential Lévy processes and has tractable returns since

$$\log\left(\frac{S_t}{S_0}\right) = \mu t + \sigma W_t + \sum_{i=1}^{N_t} Y_i. \tag{1.11}$$

Here, the rough idea is that the parameter μ controls the first moment, the parameter σ corresponds to the second moment, one-sided up-jumps ($Y_i > 0$) introduce right-skewness (third moment), one-sided down-jumps ($Y_i < 0$) left-skewness, and two-sided jumps lead to higher excess kurtosis, i.e., fourth moment. We find, that it suffices to choose a simple jump distribution with frequency $\lambda = 1$ and two equally likely fixed jump sizes

$$Y_i = \begin{cases} \log f^u & \text{with probability } 0.5, \\ \log f^d & \text{with probability } 0.5, \end{cases} \tag{1.12}$$

where $f^u \geq 1 \geq f^d$ denote the upward and downward jump factor, respectively. By construction, the occurrence of an upward jump is equivalent to multiplying the stock price by f^u. Note, that while the size of the jump is fixed by either f^u or f^d and the frequency is fixed by λ, the number of jumps per year is Poisson distributed and the timing of the jump within the year is uniformly distributed. In a given realization, several jumps may occur or none at all.

It remains to be seen how the moments of this model can be controlled. To this end, we theoretically derive the first four moments from formula (1.11) and the definitions of the moments as provided in A.1. If X is the annual return of $\mathbf{S}$ then computations (detailed in A.2) yield

$$\begin{aligned} \mathbb{E}[X] &= \mu + \frac{1}{2}\left(\log\left(f^u\right) + \log\left(f^d\right)\right) \\ \text{sd}(X) &= \sqrt{\sigma + \frac{1}{2}\left(\log\left(f^u\right)^2 + \log\left(f^d\right)^2\right)} \\ \text{skewness}(X) &= \frac{4\left(\log\left(f^u\right)^3 + \log\left(f^d\right)^3\right)}{\left(2\log\left(f^u\right)^2 + 2\log\left(f^d\right)^2 + 4\sigma\right)^{3/2}} \\ \text{kurtosis}(X) &= \Big[5\log\left(f^d\right)^4 + \left(6\log\left(f^u\right)^2 + 12\sigma\right)\log\left(f^d\right)^2 + 5\log\left(f^u\right)^4 + \\ &\quad 12\log\left(f^u\right)^2\sigma + 12\sigma^2\Big]\left(\log\left(f^u\right)^2 + \log\left(f^d\right)^2 + 2\sigma\right)^{-2}. \end{aligned} \tag{1.13}$$

As for the normal inverse Gaussian process, we fix left sides and solve this non-linear system of equations for the free parameters μ, σ, f^u, and f^d. Again, we employ Broyden's Quasi-Newton method and the double dogleg trust region method (Dennis and Schnabel, 1996). For example, if an expected return of 8%, a standard deviation of 20%, a skewness of 0, and a kurtosis of 3.5 is the choice of the return moments, then the process parameters which generate such a process are $\mu = 0.08$, $\sigma \approx 0.11$, $f^u \approx 1.18$, and $f^d \approx 0.85$. Simulating asset paths and calculating the drawdown measures is analogous to section 1.3.

We apply this model to repeat the comparative static analysis from section 1.4. Like figures 1.3

and 1.4 for the NIG model we obtain figures 1.10 and 1.11 for the jump diffusion model. The new plots are presented in A.5. For ease of comparison, the axes coincide in all aforementioned plots.

In a nutshell, the results obtained for the normal inverse Gaussian Lévy model in section 1.4 are affirmed by the results for the jump diffusion process. For the first and second moments, the plots are indistinguishable not only in relative shape but also in absolute size.

Comparing the results for the third and fourth moments, the intervals of admissible values slightly differ by construction (more details below). We find that the general trends obtained before can be verified by the jump diffusion model. There seem to be slight differences between the models for high-α CDD at high positive skewness and when the jump diffusion model approaches its highest admissible kurtosis. In summary, the general tendencies from section 1.4 can be reproduced satisfactorily.

1.5.1 Similarities and differences to the NIG model

We add a short comparison of this model to the NIG model of section 1.3 because, to allow for a sensible robustness check, the two models should not look too much alike. Some features of the two processes have to coincide, of course, for example the first four return moments which we force to coincide. Apart from that, the models appear to be quite different on theoretical grounds: the NIG Lévy process has infinite activity jumps, no diffusion and semi-heavy tails whereas the jump diffusion process has very few and fixed-height jumps, a diffusion component and no heavy tails.

Moreover, there are fewer attainable combinations of moments here than in the NIG model: For the first and second moment, there are virtually no restrictions as before. For the skewness, the set of attainable values shrinks from approximately $[-0.75, 0.75]$ in the NIG model to $[-0.50, 0.50]$ here. Whereas in the NIG model many values are possible for the kurtosis as long as they are strictly above 3, the kurtosis can be varied between 3 and 4 in this setting. We choose a kurtosis of 3.5 for the base case (instead of 4 as for the NIG model) such that we can vary the kurtosis in both directions. One way to allow more moment combinations would be to introduce more parameters, for example, different jump frequencies λ or more intricately distributed jump heights Y_i. However, an advantage of our model is that in the current parametrization its solutions are unique – just like in the NIG model.
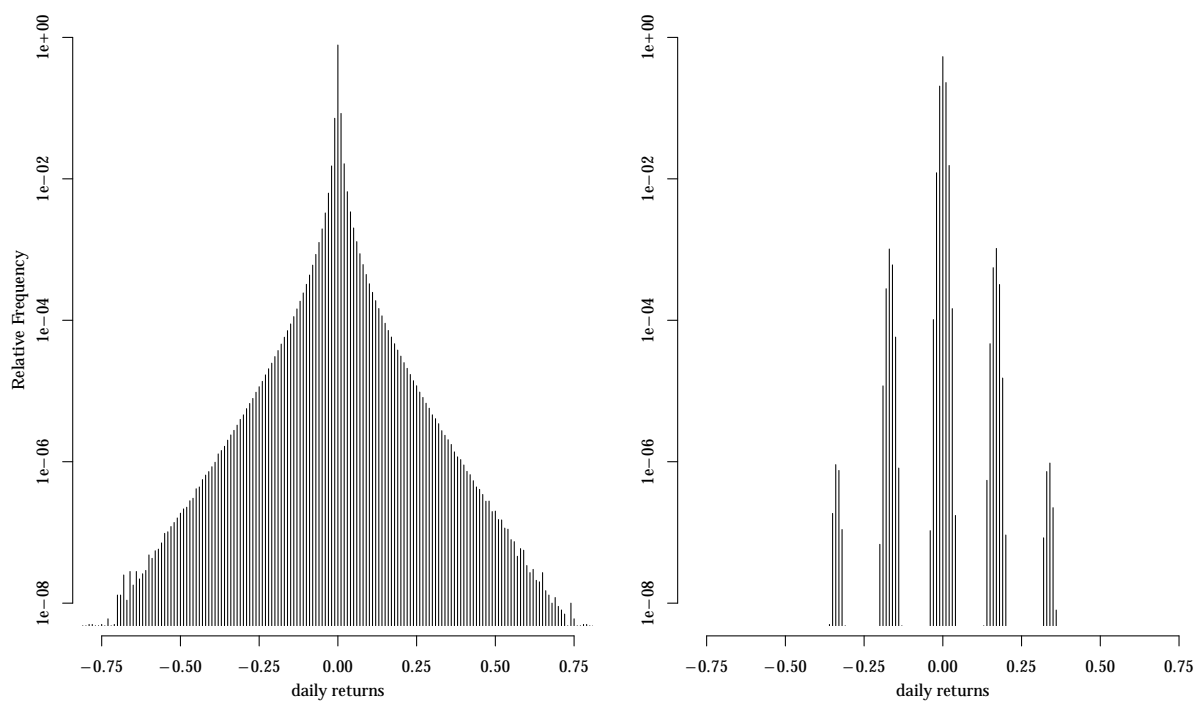
Figure 1.5: Comparison of realized daily returns from the NIG model of section 1.3 (left) and the jump diffusion model of this section. Both histograms of relative frequencies are based on one billion observations from each model for the base case of $\mu = 8\%$, $\sigma = 20\%$, skewness $= 0$ and kurtosis $= 3.5$. For better comparison, y-axes are log-scaled and the axes of both plots coincide.

Jump diffusion models and homogeneous Lévy models are somewhat similar in the sense that one may be approximated by the other (Rydberg, 1997). However, our specific jump diffusion model with fixed jump heights would be unsuitable to approximate the NIG Lévy process.[16]

The probably most straightforward way to compare both models is to generate returns from both models and compare the results, cf. the histograms in figure 1.5. For each histogram, daily returns were simulated from the base case parametrization. For better visualization, the relative frequencies are log scaled such that tail behavior can be observed more easily. Apparently, both return distributions are quite different in shape. In the right histogram, the number of jumps are clearly visible as separate peaks whereas the left histogram has a single mode. Moreover, the interval of attainable returns is more than twice as large for the NIG model. In summary, we find the jump diffusion model of this section to be sufficiently different from the NIG model to test robustness.

1.6 Drawdown Sensitivity to Shocks

In the comparative statics considered so far, one single return moment changes at a time. For example, the skewness is varied while the remaining three moments are held constant. As seen in section 1.5, Poisson jump processes can be employed to help generate processes with certain sets of moments.

However, jumps may not only constitute a modeling tool. Investors may encounter market situations in which their gut feeling tells them that a significant shock is imminent. Such additional jumps due to shocks can easily be incorporated into the model. Assuming that investors assess their moment expectations separately, the model parameters are fixed as usual. Afterwards, an additional jump component is added to the model depending on the size and direction of the suspected shock. Of course, the same is possible for the normal inverse Gaussian Lévy process from section 1.3 where a Poisson jump component can be added to the NIG component in each time interval.

---

[16]The idea in Rydberg (1997) is to approximate the infinite-activity portion of the NIG Lévy process by a Brownian motion and the larger jumps by a compound Poisson process. This is especially feasible for trying to approximate the time-continuous NIG process in more and more detail. Here, we instead sample from the process at a discrete daily frequency. Moreover, our compound Poisson process with fixed jumps is unsuitable to approximate the large jumps of the NIG process which have continuous jump heights.
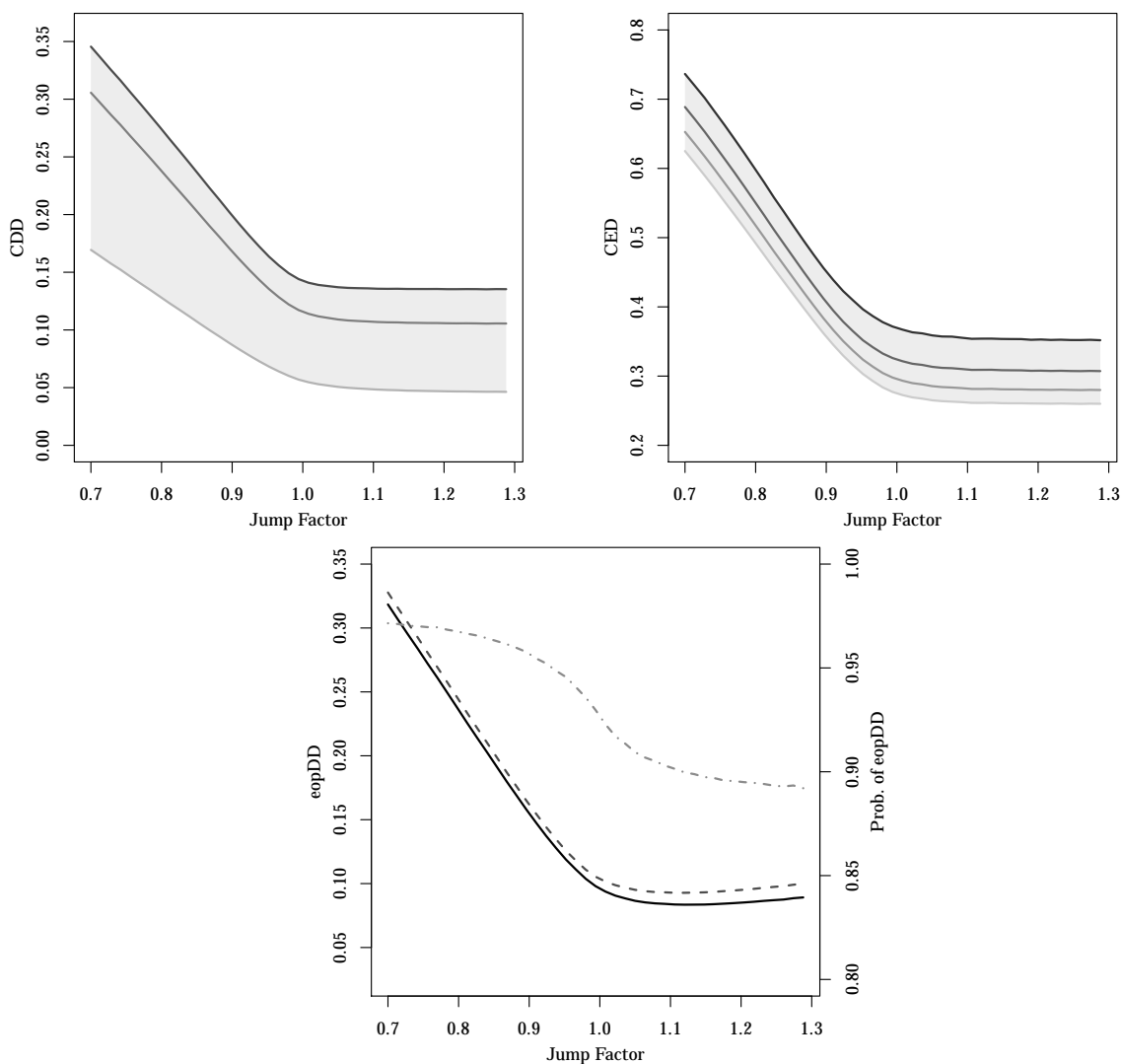
Figure 1.6: Effect on several drawdown risk measures if a shock is added to the model in its base case configuration. The size of the additional jump varies between −30% and +30%. As before, the CDD-chart (top left) contains the MDD, $CDD_{0.8}$ and ADD from top to bottom. The top right chart displays the CED for levels $0.95, 0.9, 0.85,$ and 0.8. The bottom chart contains the eopDD (bold line) and its decomposition into the probability of eopDD (dot-dashed line) and the conditional eopDD (dashed line). Each plot is based on 1,000,000 simulated paths for each jump size. All paths are simulated from the normal inverse Gaussian Lévy model with an additional Poisson jump component.

Figure 1.6 provides the simulation results for this setting. Starting from the base case, an additional jump of a size between -30% and $+30\%$ with frequency $\lambda = 1$ is implemented. As before, although jump size and jump frequency are fixed, each realization drawn from the model may or may not contain an additional jump (or many). The results show that upward jumps starkly differ from downward jumps. All drawdown measures are unaffected by upward jumps of various sizes. On the one hand, an upward jump may reduce drawdown by offsetting a previous decline. On the other hand, it may increase drawdown by setting a new benchmark high. The two effects seem to offset each other. Contrary to upward jumps, downward jumps strongly increase all measures of drawdown risk. Whereas upward jumps sometimes offset a previous downturn, downward jumps always cause drawdown by definition. Downward jumps affect all measures in an approximately linear fashion, i.e., a downward jump of -30% increases a drawdown measure by twice the absolute amount as a downward jump of -15% compared to the base case. To emphasize the size of the effect, an anticipated -30% shock in our setting may double the expected MDD or even triple expected eopDD.

Testing the validity of the results with the alternative jump diffusion model of section 1.5, we obtain the relationships displayed in figure 1.12 in A.5. The impact of the additional jump on the drawdown measures almost perfectly coincides for both models which suggests robust results.

1.7 Drawdown Sensitivity to Autocorrelation

Drawdown measures are path-dependent by nature. To derive the drawdown graph, knowledge of the path is essential. In the models considered so far, paths are constructed in a Lévy fashion, i.e., increments are independent and stationary. In the following, we relax these conditions and investigate the effect of dependence in increments on our set of drawdown measures. We employ the standard AR(1) and GARCH(1,1) models to introduce dependence in the mean and variance equations, respectively.

Within these models, we cannot hold the first four moments constant while varying the autocorrelation. Since we have witnessed the sweeping influences of the first and especially second moments in figures 1.3 and 1.10, we control for expectation and standard deviation while varying the autocorrelation.

1.7.1 Sensitivity in GARCH model

To incorporate dependence in the variance equation, often witnessed as volatility clustering, we use the Generalized Autoregressive Conditional Heteroscedasticity model (GARCH) introduced by Bollerslev (1986). To simulate series of returns $r_1, \ldots, r_T$, we choose a normal GARCH(1,1) structure given by

$$
\begin{aligned}
r_t &= C + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_t^2\right) \\
\sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{aligned}
\tag{1.14}
$$

where $C \in \mathbb{R}, \alpha, \beta \geq 0$, $\omega > 0$ and $\alpha + \beta < 1$ (Bollerslev, 1987). When fourth moments are finite[17], the 1-step autocorrelation function for the squared process is given by

$$
\rho = \mathrm{corr}\left(\varepsilon_t^2, \varepsilon_{t-1}^2\right) = \alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2},
\tag{1.15}
$$

which is monotonically increasing in α and β (Ding and Granger, 1996). To isolate the effect of autocorrelation on the drawdown measures, we prescribe an autocorrelation ρ and determine the corresponding $\alpha = \beta$. Next, we control for changes in the mean and standard deviation by setting C and ω such that the expectation and standard deviation of annual log returns are 0.08 and 0.20, respectively.[18] For each parameter set thus obtained, we simulate 1,000,000 paths and compute all drawdown measures as before.

The results are provided in figure 1.7. Unambiguously, the relationship between autocorrelation and all drawdown measures under consideration is virtually flat. Thus, volatility clustering appears to have no significant effect on drawdown measures.

1.7.2 Sensitivity in AR model

To incorporate dependence in the mean equation, we use the AR (Autoregressive) model as in Hamilton (1994). In our AR(1) model, series of daily log returns $r_1, \ldots, r_T$ are given by the difference equation

$$
r_t = C + \varphi\, r_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right),
\tag{1.16}
$$

where $r_0 = 0$, $C \in \mathbb{R}$, $|\varphi| < 1$ and $\sigma_\varepsilon^2 > 0$. The annual return is given by $R = \sum_{t=1}^{T=252} r_t$.

---

[17]For the GARCH(1,1), $\mathbb{E}\left[\varepsilon_t^4\right] < \infty$ if $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$, which we require in the following (Bollerslev, 1987).

[18]To this end, we initialize the GARCH(1,1) process for the daily returns in the stationary mean and variance, which are given by C and $\frac{\omega}{1-\alpha-\beta}$, respectively.
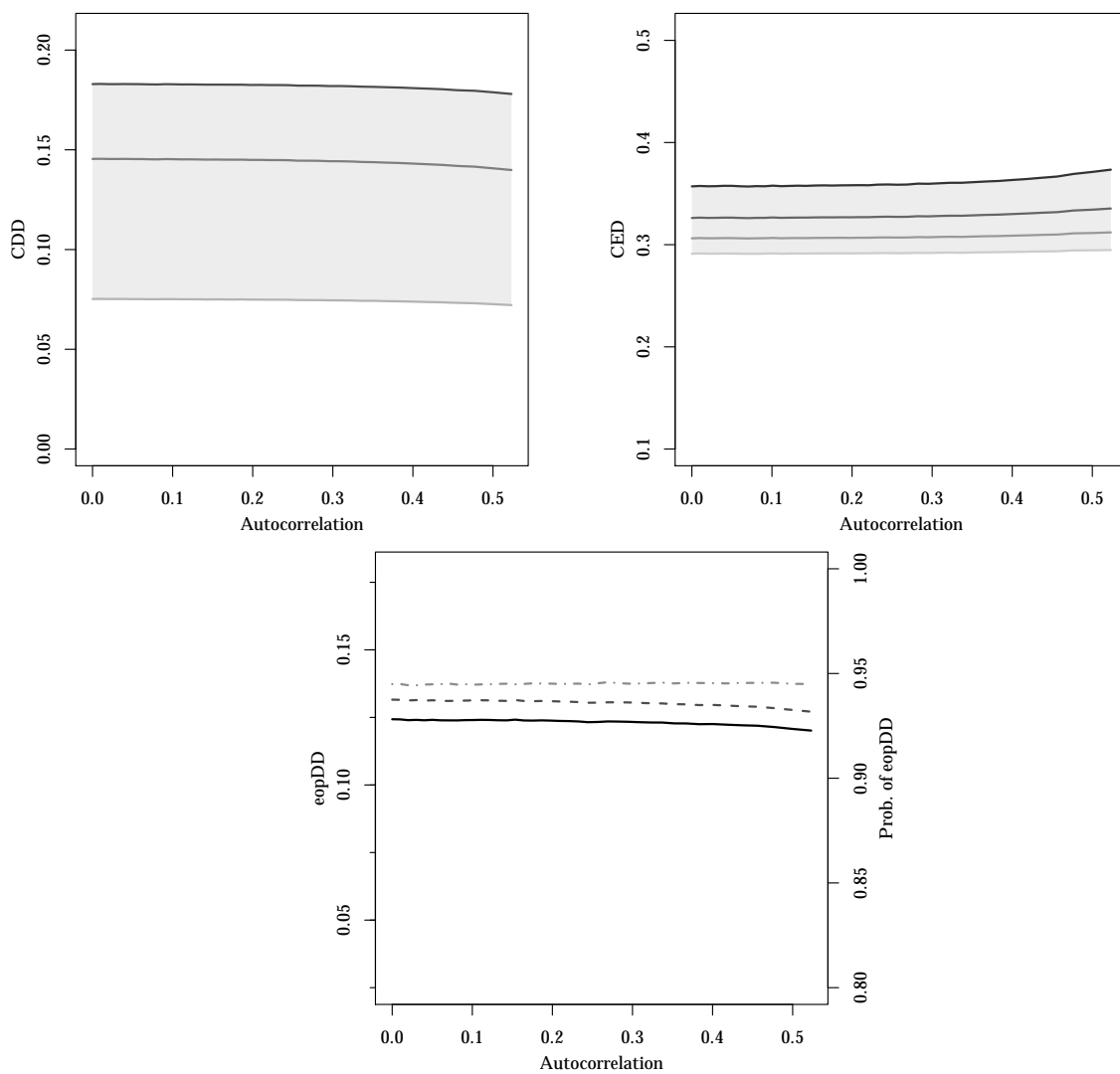
Figure 1.7: Effect of autocorrelation on several drawdown risk measures in a GARCH(1,1) model. Mean and standard deviation of annual returns are held constant. As before, the CDD-chart (top left) contains the MDD, $CDD_{0.8}$ and ADD from top to bottom. The top right chart displays the CED for levels $0.95, 0.9, 0.85$, and 0.8. The bottom chart contains the eopDD (bold line) and its decomposition into the probability of eopDD (dot-dashed line) and the conditional eopDD (dashed line). Each plot is based on 1,000,000 simulated paths for each fixed autocorrelation.

For the AR(1) process, we analyze how the drawdown measures are affected by autocorrelation in the returns. The latter is immediately available as

$$\text{corr}\,(r_t, r_{t-1}) = \varphi \tag{1.17}$$

according to Hamilton (1994). Then, we determine

$$\mathbb{E}[R] = \frac{CT}{1-\varphi} \qquad \text{and} \qquad \mathbb{V}\text{ar}(R) = \sigma_\varepsilon^2 \, \Xi(\varphi, T) \tag{1.18}$$

where Ξ is a function of φ and T as detailed and derived in appendix A.3. We start with the base case in which $\mathbb{E}[R] = \mu = 8\%$, $\sqrt{\mathbb{V}\text{ar}(R)} = \sigma = 20\%$. Next, we set φ to model autocorrelation and choose C and σ_ε^2 such that the annual expectation and standard deviation remain constant, i.e.,

$$C = \frac{\mu(1-\varphi)}{252} \qquad \text{and} \qquad \sigma_\varepsilon^2 = \frac{\sigma^2}{\Xi(\varphi, 252)}. \tag{1.19}$$

Again, we compute the results for all drawdown measures of section 1.2. The plots contained in figure 1.8 are similar for all drawdown measures: autocorrelation in the returns has close to no impact. In general, a slight downward trend can be observed, i.e., the higher the autocorrelation, the slightly smaller the drawdown measures.

One of the drawdown measures under consideration, the CED, has been studied with regards to AR(1) autocorrelation before. Goldberg and Mahmoud (2017) report a significantly positive relationship between the AR-parameter φ and the CED. Although these results seem to contradict our findings, there is a plausible explanation. In the AR model without any corrections, increasing the autocorrelation φ also drives up the return variance. This is intuitive and validated by

$$\mathbb{V}\text{ar}\,(r_t) = \frac{\sigma_\varepsilon^2}{1-\varphi^2}$$

as in formula (A.2) in A.3 because by increasing φ the denominator becomes small which increases the variance. Hence, as Goldberg and Mahmoud (2017) increase φ and intend to measure the effect of autocorrelation, they accidentally also measure the effect of variance. From our previous results in figures 1.3 and 1.10, we know that changes in variance have a large impact on all drawdown measures. Hence, we conjecture that autocorrelation in itself does not increase drawdown but that increasing the autocorrelation parameter in the model increases variance which we have seen, in turn, to increase drawdown.

To be sure, we compare the effect of autocorrelation on the CED while applying different controls. Figure 1.9 contains the results. First, we increase autocorrelation but do not control for anything.
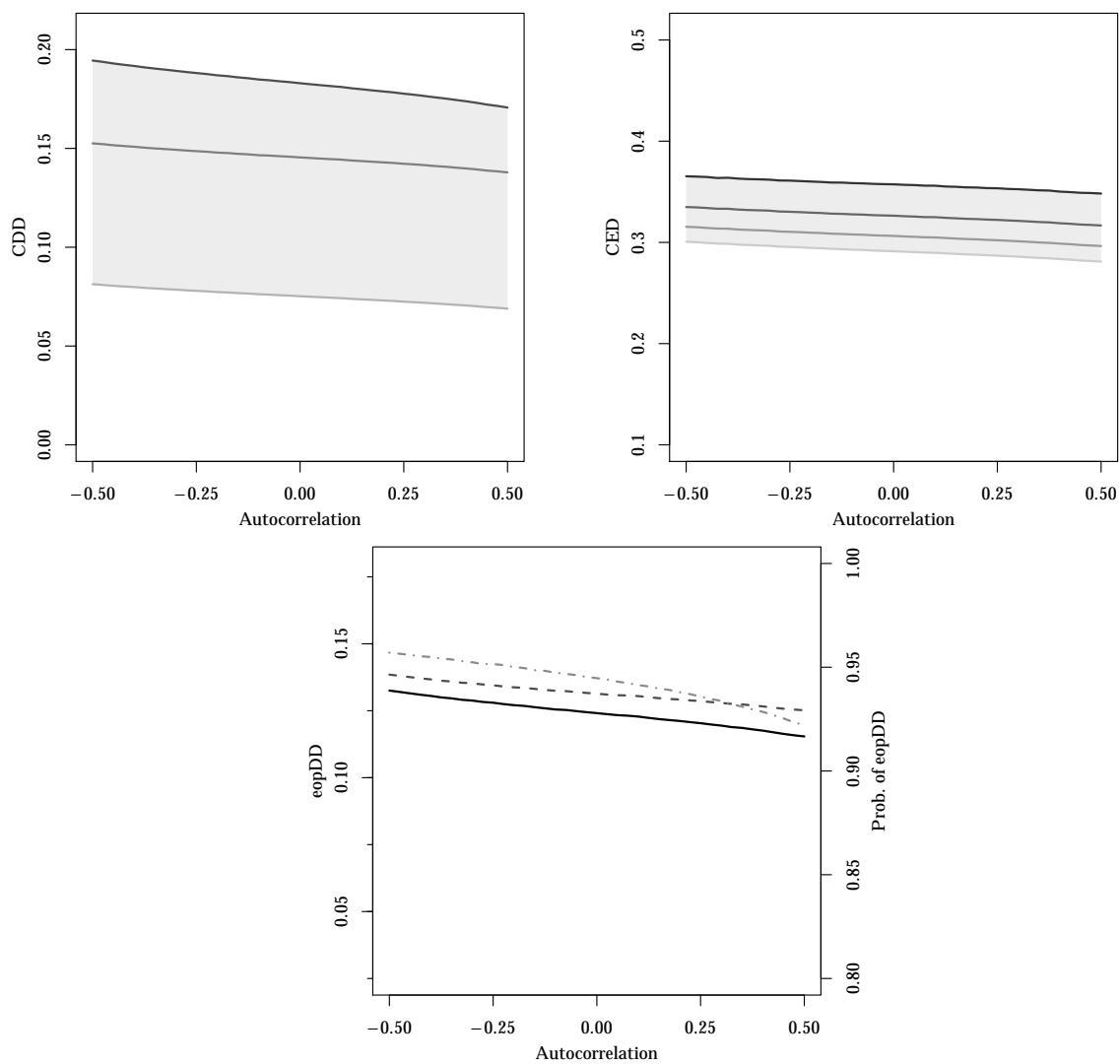
Figure 1.8: Effect of autocorrelation on several drawdown risk measures in an AR(1) model. Mean and standard deviation of annual returns are held constant. As before, the CDD-chart (top left) contains the MDD, $CDD_{0.8}$ and ADD from top to bottom. The top right chart displays the CED for levels $0.95, 0.9, 0.85$, and 0.8. The bottom chart contains the eopDD (bold line) and its decomposition into the probability of eopDD (dot-dashed line) and the conditional eopDD (dashed line). Each plot is based on 1,000,000 simulated paths for each fixed autocorrelation.
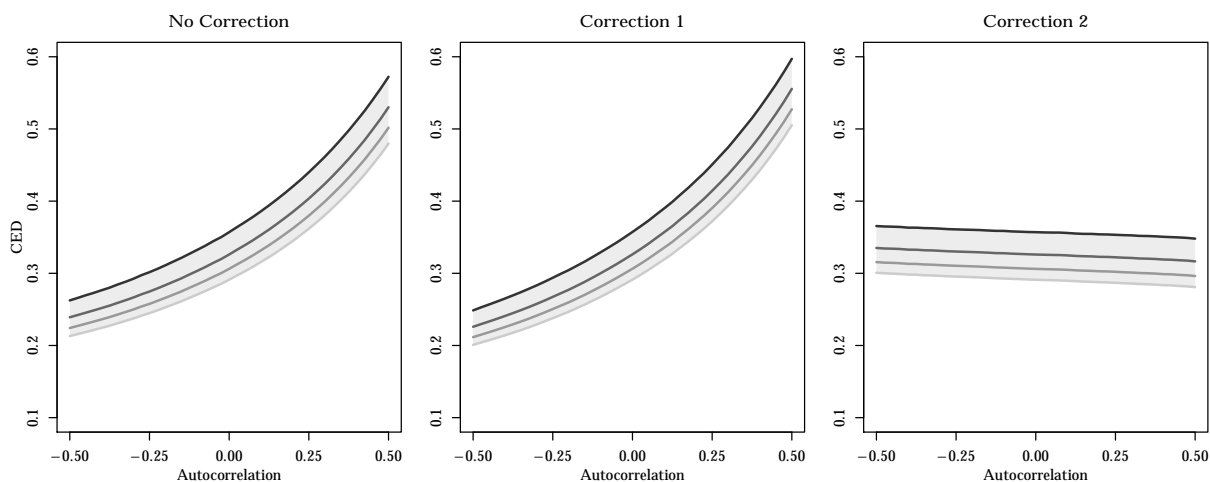
Figure 1.9: Is the increase in CED with AR(1) autocorrelation reported in Goldberg and Mahmoud (2017) driven by autocorrelation or variance? All three plots are generated from the AR(1) model. In the left plot, φ is increased with no correction for simultaneous changes in expectation and variance. Correction 1 holds the expectation constant as φ changes. In the right plot, we account for both expectation and variance, thus isolating the effect of autocorrelation best. In all plots, the CED is presented for levels $0.95, 0.9, 0.85$, and 0.8. Each plot is based on 1,000,000 simulated paths for each fixed autocorrelation.

As expected, the CED strongly increases, which is the same result as in Goldberg and Mahmoud (2017). Next, we hold the expectation constant while varying φ. The effect on the CED remains strong as before. Third, we also correct for the variance and observe that the effect on the CED disappears entirely. Repeating the analysis for the other drawdown measures yields similar results. Thus, we conclude that it is variance that drives up the drawdown measures and AR(1) autocorrelation has no significant effects.

Returning to the GARCH(1,1) model, we can identify the same phenomenon. If we stop controlling for mean and variance, all drawdown measures increase. Again, we observe that in order to increase autocorrelation, we increase α and β which, in turn, increases the unconditional variance $\frac{\omega}{1-\alpha-\beta}$. We conclude that in both models it is important to control for the variance to identify the true impact of autocorrelation which is in both cases negligible.

1.8 Conclusion

This paper's central contribution to the study of drawdown risk measures is twofold. First, it adds to the motivation of drawdown measures a psychological component which is founded in regret theory and behavioral finance. A new end-of-period drawdown measure which captures this psychological component is introduced and analyzed theoretically. Second, this paper relates

the study of higher moments of returns to the field of drawdown risk measures. It thus provides answers to natural questions like how asymmetry and fat tails influence how risky a strategy is deemed by different drawdown risk measures.

The new eopDD measure makes it possible to take into account that investors often encounter fixed investment horizons, evaluation periods or portfolio rebalancing cycles such that emphasizing the performance at the end of the time horizon is important. The new measure has straightforward interpretations in terms of regret with respect to optimal market timing and satisfies the desirable properties of generalized deviation measures (Goldberg and Mahmoud, 2017). Empirically, it is found to rank portfolios differently than previous drawdown measures.

To analyze the link between return moments and drawdown measures, analytic results for the drawdown quantities would be ideal. However, explicit results for drawdown quantities are almost non-existent in the literature. Certainly, an analytic study of drawdown quantities at least requires results about the suprema of the underlying stochastic processes, obtaining which is in itself far from trivial. For the basic drawdown graph, most existing results are either stated specifically for the (drifted) Brownian motion process (cf. Landriault et al. (2015) and the references therein), or require the stochastic process to be spectrally negative (e.g., Mijatović and Pistorius (2012)) – or spectrally positive in the case of drawup instead of drawdown (e.g., Pistorius (2004)) – which by definition means that the Lévy measure is supported by $(-\infty, 0)$ (Doney, 2007). Intuitively, spectrally negative processes may move upwards and downwards by diffusion, may jump downwards, but may never jumps upwards (Doney, 2007). Unfortunately, neither the NIG Lévy model nor our jump diffusion model are spectrally negative. The NIG model has upward jumps as its Lévy measure is not concentrated on $(-\infty, 0)$, compare, for example, the Lévy-Khintchine formula in Rydberg (1997). In our jump diffusion model, the compound Poisson process explicitly jumps upwards with probability $\frac{1}{2}$.

Moreover, results are usually only available for the Maximum Drawdown because it can be derived from the stopping time τ_a of the drawdown process first crossing the threshold a (Landriault et al., 2017b). Mostly, other quantities (related to this stopping time) of the drawdown graph are considered, for example, the last time the price was at its supremum prior to τ_a, the running maximum and minimum at τ_a, the overshoot and undershoot of the drawdown process at τ_a (cf. the sextuple law of Mijatović and Pistorius (2012) or the results in Landriault et al. (2017a) both formulated in terms of joint Laplace transforms).

In fact, it is already difficult to obtain distributional results for the running maximum of a Lévy

process, see the extensive references in Kwásnicki et al. (2013). Although the general case has been addressed as early as in Baxter and Donsker (1957), explicit results have yet been found for the Brownian motion, the Cauchy process, the compound Poisson with Lévy Khintchine exponent $1 - \cos \xi$ and the (reversed) Poisson process with drift – but neither for the NIG nor for the jump diffusion processes (Kwásnicki et al., 2013).

In light of the difficulties with an analytic approach, which we leave for further research, we instead analyze the link between return moments and drawdown measures in a comprehensive simulation study. Our simulation study includes the new end-of-period drawdown measure as well as all standard drawdown measures covered in the literature. We employ a normal inverse Gaussian Lévy model for simulating the paths because it allows wide-ranging manipulation and analytic computation of the return moments which is vital for this kind of analysis. We double-check the results with a second tailor-made jump diffusion model and obtain confirming results.

The simulation study finds that for the first and second moments all drawdown measures react similarly and in line with standard moment preferences that high odd moments are desirable and high even moments are not. For higher moments, some results are rather counter-intuitive. For example, high kurtosis is rewarded by the MDD but penalized by the $\text{CED}_{0.95}$. Consequently, practitioners are advised to scrutinize if their choice of risk measure is consistent with their needs. For skewness and kurtosis, members of the CED family generally yield more promising results than members of the CDD or eopDD family. For high confidence levels, the CED complies with moment preferences for skewness and kurtosis while the ADD does neither.

Extending the model to include additional jumps, we find that upward shocks have close to no impact whereas downward shocks raise all drawdown measures significantly. Finally, we investigate the drawdown measures' sensitivity to autocorrelation and show that effects are minute compared to the moment sensitivities.

A Appendix

A.1 Definition of Moments

Many essential properties of probability distributions are encoded in the distributions' moments. In general, the n-th moment of a random variable refers to the expectation of its n-th power. It is commonplace to make higher moments of different distributions comparable by centralization

around their mean and normalization with their standard deviation. This way, the (centralized and normalized) moments equal the well-known quantities expectation, variance, skewness, and kurtosis. For some random variable X, for example, the end-of-period returns of a stochastic process, the definitions of the moments are as in the following table.

Name	Moment	Definition
Expectation	1st Moment	$\mathbb{E}\left[X\right]$
Std. dev.	$\sqrt{\text{2nd Centr. Moment}}$	$\sigma_X = \sqrt{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}$
Skewness	3rd Centr. Norm. Moment	$\mathbb{E}\left[\left(\dfrac{X - \mathbb{E}[X]}{\sigma_X}\right)^3\right]$
Kurtosis	4th Centr. Norm. Moment	$\dfrac{\mathbb{E}\left[(X - \mathbb{E}[X])^4\right]}{\sigma_X^4}$

Table 1.3: Definitions of (higher) moments.

A.2 Moment Formulas for the Jump Diffusion Process

In the description of the jump diffusion model in section 1.5, the derivation of the system of equations (1.13) for the first four moments has been omitted. In the following, we provide a detailed guide for computing the kurtosis since this constitutes the most complicated case. For brevity, we will however not explicitly state all formulas. For illustration, we will in the end provide the explicit formula for the first moment. Recall formula (1.11) for the log return of the price process for one year

$$X := \mu + \sigma W_1 + \sum_{i=1}^{N_1} Y_i,$$

where W_t is the Wiener process at time t and $N_t \sim \text{Poi}(\lambda t)$ with $\lambda = 1$. Recall from the previous section that

$$\text{kurtosis}(X) = \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^4\right]}{\mathbb{V}\text{ar}(X)^2}.$$

In the numerator, we apply binomial formulas, use linearity of the expectation, and collect terms to obtain

$$\mathbb{E}[X^4] - 4\,\mathbb{E}X^3]\mathbb{E}[X] + 6\,\mathbb{E}[X^2]\mathbb{E}[X]^2 - 3\,\mathbb{E}[X]^4.$$

In the denominator, we use that

$$\mathbb{V}\text{ar}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Hence, it remains to compute $\mathbb{E}[X^k]$ for $k = 1, \ldots, 4$ and collect terms. We rewrite X as $X = Z + P$ where $Z = \mu + \sigma W_1 \sim \mathcal{N}(\mu, \sigma)$ and $P = \sum_{i=1}^{N_1} Y_i$. Once again by binomial formulas, linearity of the expectation and independence of Z and P, we can compute $\mathbb{E}[X^k]$ once all the expectations $\mathbb{E}[Z^k]$ and $\mathbb{E}[P^k]$, $k = 1, \ldots, 4$ are known. To this end, we use the moment-generating functions of Z and P.[19] For $Z \sim \mathcal{N}(\mu, \sigma)$, the moment-generating function is given by

$$M_Z(x) = \mathbb{E}\left[\exp(x \cdot Z)\right] = \exp\left(\mu x + \frac{1}{2}\sigma^2 x^2\right).$$

For the compound Poisson process P, the moment-generating function is

$$M_P(x) = \exp\left(M_{Y_1}(x) - 1\right),$$

where $M_{Y_1}(x)$ is the moment-generating function of Y_1. As is well-known, $\mathbb{E}\left[X^n\right]$ equals the n-th derivative of $M_X(x)$ evaluated at $x = 0$. This procedure yields

$$\mathbb{E}[Z] = \mu \qquad\qquad \mathbb{E}[Z^2] = \mu^2 + \sigma$$

$$\mathbb{E}[Z^3] = \mu^3 + 3\mu\sigma \qquad\qquad \mathbb{E}[Z^4] = \mu^4 + 6\mu^2\sigma + 3\sigma^2.$$

For P, we first need the moment-generating function $M_{Y_1}(x)$ of

$$Y_1 = \begin{cases} \log f^u & \text{with probability } 0.5, \\ \log f^d & \text{with probability } 0.5, \end{cases}$$

which is $\mathbb{E}[\exp(x \cdot Y_1)] = \frac{1}{2}\exp\left(x \log f^u\right) + \frac{1}{2}\exp\left(x \log f^d\right)$. Thus,

$$M_P(x) = \exp\left(\frac{1}{2}(f^u)^x + \frac{1}{2}\left(f^d\right)^x - 1\right).$$

In total, it now remains to compute well-behaved derivatives which is straightforward and thus omitted. Then, the formula for the kurtosis from the system of equations (1.13) is obtained. As an example, we now derive the explicit formula for the first moment:

$$\frac{\mathrm{d}M_P(x)}{\mathrm{d}x} = \left(\frac{1}{2}(f^u)^x \log(f^u) + \frac{1}{2}\left(f^d\right)^x \log\left(f^d\right)\right) \exp\left(\frac{1}{2}(f^u)^x + \frac{1}{2}\left(f^d\right)^x - 1\right)$$

and evaluating at $x = 0$ yields

$$\left.\frac{\mathrm{d}M_P(x)}{\mathrm{d}x}\right|_{x=0} = \frac{1}{2}\log(f^u) + \frac{1}{2}\log\left(f^d\right) = \mathbb{E}[P].$$

In total, we obtain for the expectation of the log return

$$\mathbb{E}[X] = \mathbb{E}[Z] + \mathbb{E}[P] = \mu + \frac{1}{2}\log(f^u) + \frac{1}{2}\log\left(f^d\right)$$

as provided in the system of equations (1.13).

---

[19]The two postulated formulas for the moment-generating functions can be verified by straightforward computations from the respective definitions.

A.3 Moment formulas for the AR(1) process

For the AR(1) model applied in section 1.7, we derive the expectation and variance of annual returns R. To recall, series of daily returns $r_1, \ldots, r_T$ are given in the AR(1) model by the difference equation

$$r_t = C + \varphi \, r_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right),$$

where $r_0 = 0$, $C \in \mathbb{R}$, $|\varphi| < 1$ and $\sigma_\varepsilon^2 > 0$. In the following, we use three results from Hamilton (1994):

$$\mathbb{E}\left[r_t\right] = \nu = \frac{C}{1 - \varphi} \tag{A.1}$$

$$\mathbb{V}\text{ar}\left(r_t\right) = \frac{\sigma_\varepsilon^2}{1 - \varphi^2} \tag{A.2}$$

$$\mathbb{E}\left[r_{t+n} \cdot r_t\right] = \nu^2 + \frac{\sigma_\varepsilon^2}{1 - \varphi^2}\varphi^n. \tag{A.3}$$

For the annual return $R = \sum_{t=1}^{T=252} r_t$, we compute

$$\mathbb{E}\left[R\right] = \mathbb{E}\left[\sum_{t=1}^{T} r_t\right] = \sum_{t=1}^{T} \mathbb{E}[r_t] \overset{(A.1)}{=} \sum_{t=1}^{T} \frac{C}{1 - \varphi} = \frac{CT}{1 - \varphi}.$$

For the variance, we obtain

$$\mathbb{V}\text{ar}(R) = \mathbb{V}\text{ar}\left(\sum_{t=1}^{T} r_t\right) = \sum_{t=1}^{T} \mathbb{V}\text{ar}\left(r_t\right) + 2 \sum_{1 \leq i < j \leq T} \text{Cov}\left(r_i, r_j\right)$$

$$\overset{(A.2)}{=} \sum_{t=1}^{T} \frac{\sigma_\varepsilon^2}{1 - \varphi^2} + 2 \sum_{1 \leq i < j \leq T} \left(\mathbb{E}\left[r_i r_j\right] - \mathbb{E}\left[r_i\right]\mathbb{E}\left[r_j\right]\right)$$

$$\overset{(A.1)}{=} \frac{\sigma_\varepsilon^2 T}{1 - \varphi^2} + 2 \sum_{1 \leq i < j \leq T} \mathbb{E}\left[r_i r_j\right] - 2\frac{T^2 - T}{2}\left(\frac{C}{1 - \varphi}\right)^2.$$

To apply equation (A.3) to the above sum, we count the number of n-distant pairs which is $T - n$ for $n = 1, \ldots, T$. Thus,

$$\mathbb{V}\text{ar}(R) = \frac{\sigma_\varepsilon^2 T}{1 - \varphi^2} + 2 \sum_{n=1}^{T} (T - n)\left(\nu^2 + \frac{\sigma_\varepsilon^2}{1 - \varphi^2}\varphi^n\right) - \frac{C^2 T(T-1)}{(1 - \varphi)^2}$$

$$= \frac{\sigma_\varepsilon^2 T}{1 - \varphi^2} + 2\nu^2 \sum_{n=1}^{T} (T - n) + \frac{2\sigma_\varepsilon^2}{1 - \varphi^2} \sum_{n=1}^{T} (T - n)\varphi^n - \frac{C^2 T(T-1)}{(1 - \varphi)^2}$$

$$= \frac{\sigma_\varepsilon^2 T}{1 - \varphi^2} + \nu^2 T(T-1) - \frac{2\sigma_\varepsilon^2 \varphi\left(-\varphi^T + \varphi T - T + 1\right)}{(1 - \varphi^2)(1 - \varphi)^2} - \frac{C^2 T(T-1)}{(1 - \varphi)^2}$$

$$\overset{(A.1)}{=} \frac{\sigma_\varepsilon^2 T(1 - \varphi)^2 - 2\sigma_\varepsilon^2 \varphi\left(-\varphi^T + \varphi T - T + 1\right)}{(1 - \varphi^2)(1 - \varphi)^2}$$

$$= \sigma_\varepsilon^2 \cdot \underbrace{\left(\frac{T(1 - \varphi)^2 - 2\varphi(-\varphi^T + \varphi T - T + 1)}{(1 - \varphi^2)(1 - \varphi)^2}\right)}_{=:\Xi(\varphi, T)}.$$

A.4 Proof of Uniqueness of the NIG Model

We prove that the system of equations (1.9) from section 1.3 has at most one set of solutions $(\alpha, \beta, \mu, \delta)$. The proof idea is to assume two sets of parameters $\theta_1 = (\alpha_1, \beta_1, \mu_1, \delta_1)$ and $\theta_2 = (\alpha_2, \beta_2, \mu_2, \delta_2)$ which yield the same moments and show that θ_1 and θ_2 must coincide. In other words, we assume

$$\mu_1 + \frac{\delta_1 \beta_1}{\gamma_1} = \mu_2 + \frac{\delta_2 \beta_2}{\gamma_2} \tag{A.4}$$

$$\sqrt{\frac{\delta_1 \alpha_1^2}{\gamma_1^3}} = \sqrt{\frac{\delta_2 \alpha_2^2}{\gamma_2^3}} \tag{A.5}$$

$$\frac{3\beta_1}{\alpha_1 \sqrt{\delta_1 \gamma_1}} = \frac{3\beta_2}{\alpha_2 \sqrt{\delta_2 \gamma_2}} \tag{A.6}$$

$$3 + 3\left(1 + 4\frac{\beta_1^2}{\alpha_1^2}\right)\frac{1}{\delta_1 \gamma_1} = 3 + 3\left(1 + 4\frac{\beta_2^2}{\alpha_2^2}\right)\frac{1}{\delta_2 \gamma_2}, \tag{A.7}$$

where $\gamma_i = \sqrt{\alpha_i^2 - \beta_i^2}$, $0 \leq |\beta_i| < \alpha_i$, $\delta_i > 0$ and $\mu_i \in \mathbb{R}$ for $i = 1, 2$ and possibly different sets of parameters $\theta_1 \neq \theta_2$.

From (A.6), we obtain

$$\frac{\beta_1}{\alpha_1 \sqrt{\delta_1 \gamma_1}} = \frac{\beta_2}{\alpha_2 \sqrt{\delta_2 \gamma_2}} \iff \frac{\beta_1^2}{\alpha_1^2 \delta_1 \gamma_1} = \frac{\beta_2^2}{\alpha_2^2 \delta_2 \gamma_2}, \tag{A.8}$$

since β_1 and β_2 have the same sign because the denominator is strictly positive in the first equation due to the parameter restrictions of the model.

From (A.7), we obtain

$$\left(1 + 4\frac{\beta_1^2}{\alpha_1^2}\right)\frac{1}{\delta_1 \gamma_1} = \left(1 + 4\frac{\beta_2^2}{\alpha_2^2}\right)\frac{1}{\delta_2 \gamma_2} \iff \frac{1}{\delta_1 \gamma_1} + 4\frac{\beta_1^2}{\alpha_1^2 \delta_1 \gamma_1} = \frac{1}{\delta_2 \gamma_2} + 4\frac{\beta_2^2}{\alpha_2^2 \delta_2 \gamma_2}$$

and using (A.8) we get

$$\delta_1 \gamma_1 = \delta_2 \gamma_2. \tag{A.9}$$

Inserting (A.9) into (A.8), we obtain

$$\frac{\beta_1^2}{\alpha_1^2} = \frac{\beta_2^2}{\alpha_2^2} \quad \text{and} \quad \frac{\beta_1}{\alpha_1} = \frac{\beta_2}{\alpha_2} \tag{A.10}$$

by the same sign argument for β as above and $\alpha > 0$. Using (A.5) in combination with (A.9), we obtain

$$\frac{\delta_1 \alpha_1^2}{\gamma_1^3} = \frac{\delta_2 \alpha_2^2}{\gamma_2^3} \iff \frac{\gamma_1 \delta_1 \alpha_1^2}{\gamma_1^4} = \frac{\gamma_2 \delta_2 \alpha_2^2}{\gamma_2^4} \iff \frac{\alpha_1^2}{\gamma_1^4} = \frac{\alpha_2^2}{\gamma_2^4},$$

which together with (A.10) yields

$$\frac{\alpha_1^2 \beta_1^2}{\gamma_1^4 \beta_1^2} = \frac{\alpha_2^2 \beta_2^2}{\gamma_2^4 \beta_2^2} \iff \frac{\beta_1^2}{\gamma_1^4} = \frac{\beta_2^2}{\gamma_2^4} \iff \frac{\beta_1}{\gamma_1^2} = \frac{\beta_2}{\gamma_2^2}. \tag{A.11}$$

Applying the definition of γ, we obtain

$$\begin{aligned} \frac{\gamma_1^2}{\beta_1} = \frac{\gamma_2^2}{\beta_2} &\iff \frac{\alpha_1^2 - \beta_1^2}{\beta_1} = \frac{\alpha_2^2 - \beta_2^2}{\beta_2} \\ &\iff \frac{\beta_1 \left(\alpha_1^2 - \beta_1^2 \right)}{\beta_1^2} = \frac{\beta_2 \left(\alpha_2^2 - \beta_2^2 \right)}{\beta_2^2} \\ &\iff \beta_1 \left(\frac{\alpha_1^2}{\beta_1^2} - 1 \right) = \beta_2 \left(\frac{\alpha_2^2}{\beta_2^2} - 1 \right) \\ &\iff \beta_1 = \beta_2, \end{aligned}$$

where the last equivalence is a direct consequence of (A.10). Immediately, we obtain

$$\alpha_1 = \alpha_2 \qquad \text{and} \qquad \gamma_1 = \gamma_2$$

by (A.10) and by definition, respectively. From (A.9), we obtain

$$\delta_1 = \delta_2 \qquad \text{and} \qquad \mu_1 = \mu_2$$

from (A.4). Hence, we conclude that the set of parameters $(\alpha, \beta, \mu, \delta)$ which solves the system of equations 1.9 is unique.

A.5 Simulation Results for the Jump Diffusion Process

This appendix provides details to the results summarized in section 1.5 (figures 1.10 and 1.11) and in section 1.6 (figure 1.12).
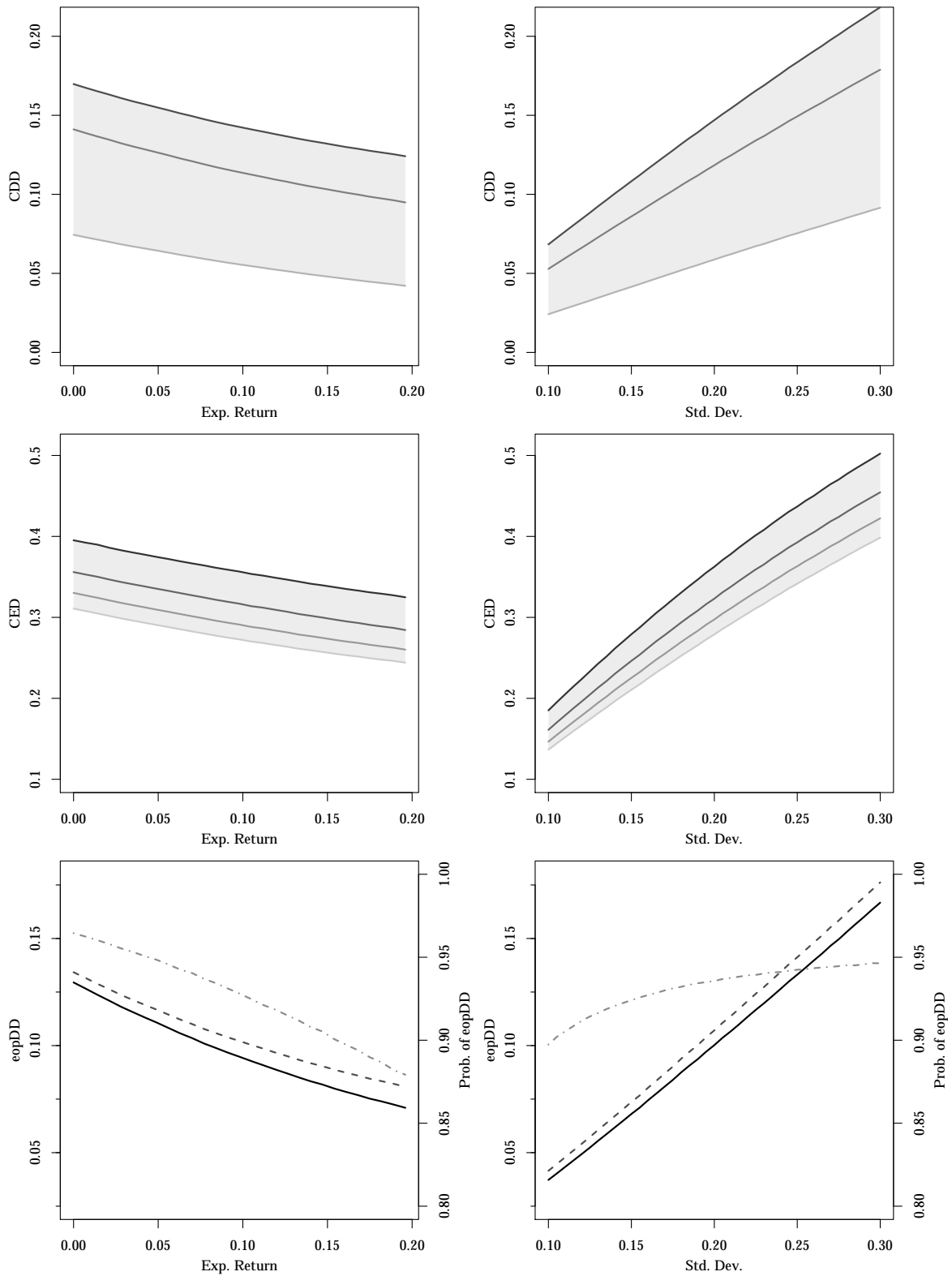
Figure 1.10: Comparative static for 1st moment (left column) and 2nd moment. The CDD-charts (first row) contain the MDD, $CDD_{0.8}$ and ADD from top to bottom. In the middle row, we display the CED for levels $0.95, 0.9, 0.85,$ and 0.8. The bottom row contains the eopDD (bold line) and its decomposition into the probability of eopDD (dot-dashed line) and the conditional eopDD (dashed line). All paths are simulated from the jump diffusion process introduced in section 1.5.
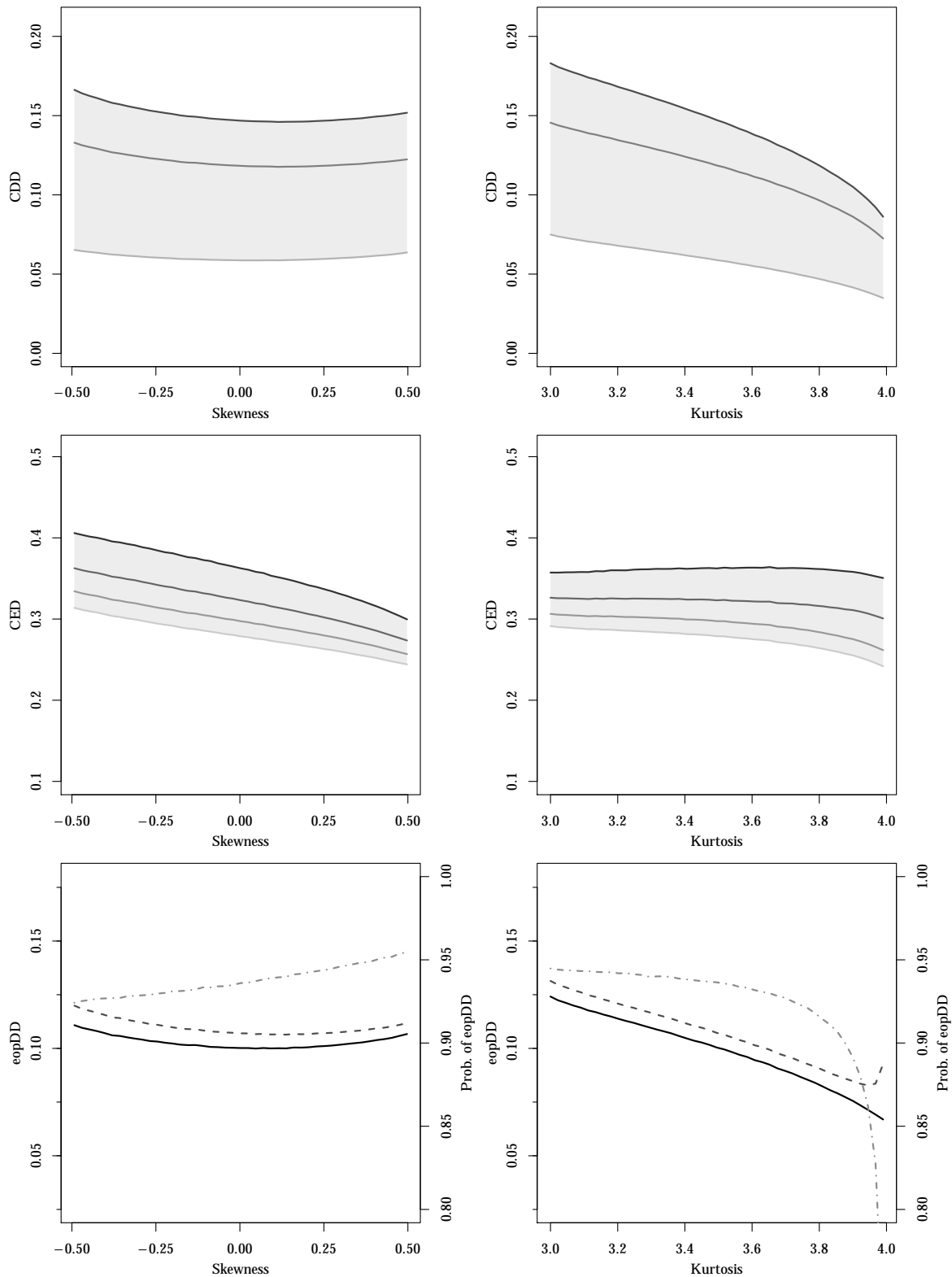
Figure 1.11: Comparative static for 3rd moment (left column) and 4th moment. The CDD-charts (first row) contain the MDD, $CDD_{0.8}$ and ADD from top to bottom. In the middle row, we display the CED for levels $0.95, 0.9, 0.85$, and 0.8. The bottom row contains the eopDD (bold line) and its decomposition into the probability of eopDD (dot-dashed line) and the conditional eopDD (dashed line). All paths are simulated from the jump diffusion process introduced in section 1.5.
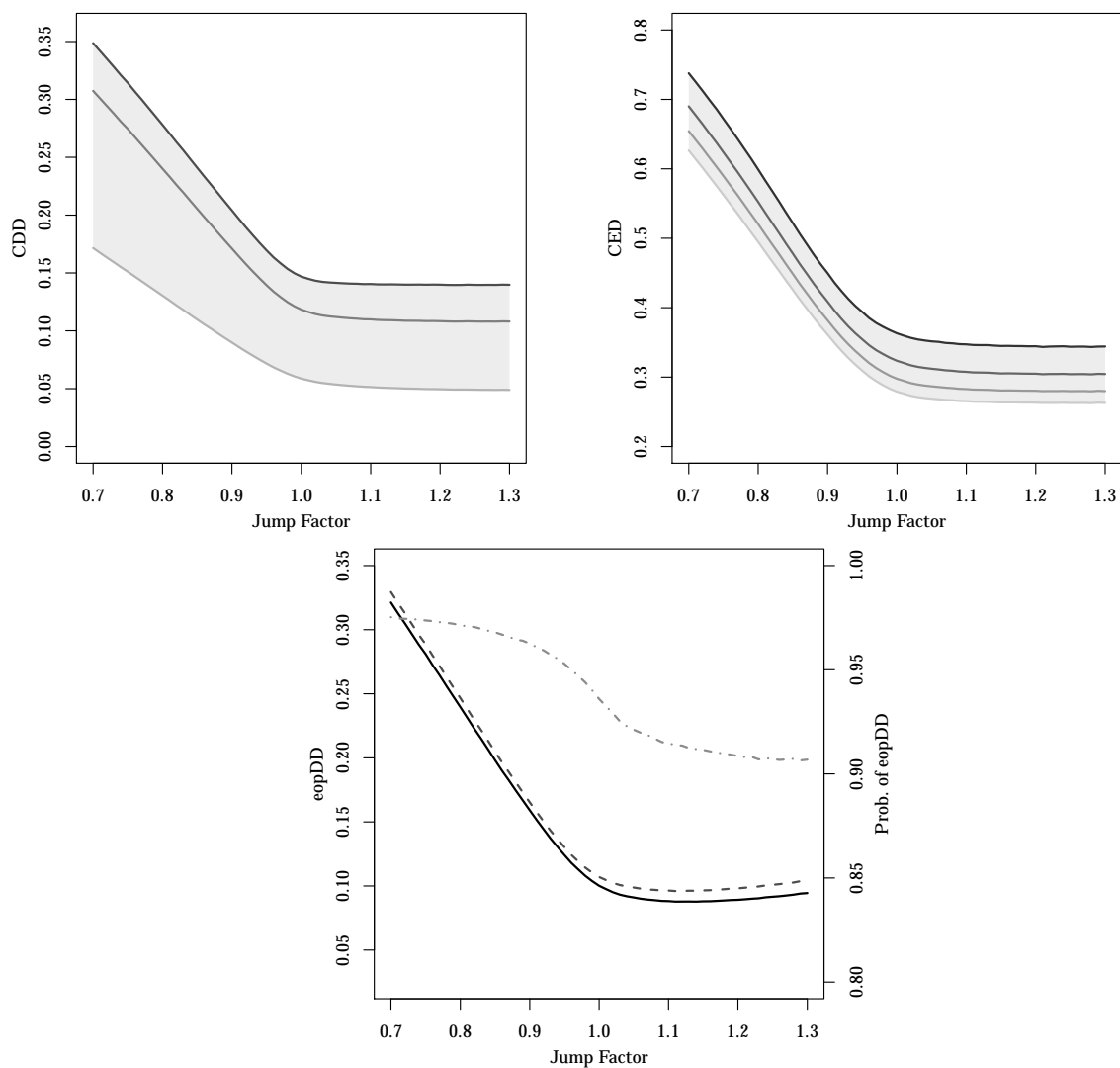
Figure 1.12: Effect on several drawdown risk measures if a shock is added to the model in its base case configuration. The size of the additional jump varies between -30% and $+30\%$. As before, the CDD-chart (top left) contains the MDD, $CDD_{0.8}$ and ADD from top to bottom. The top right chart displays the CED for levels $0.95, 0.9, 0.85$, and 0.8. The bottom chart contains the eopDD (bold line) and its decomposition into the probability of eopDD (dot-dashed line) and the conditional eopDD (dashed line). All paths are simulated from the jump diffusion process introduced in section 1.5.

2 Drawdown Measures: Are They All the Same?

Joint work with Olaf Korn and Christian Schwehm.

Accepted for publication in the Journal of Portfolio Management.[1]

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.

2.1 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 (Kahneman and Riepe, 1998; 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 and academia[2], 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). Among these, the maximum drawdown is undoubtedly the most well-known measure. However, as new drawdown-based measures have emerged, it is important to gain a better understanding of their potential. 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? How could these measures be further improved to fit investor preferences? Is the maximum drawdown still the most reasonable choice or should other drawdown measures be preferred? 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

---

[2]The drawdown concept was first floated and propagated by finance practitioners, such as Young (1991), Burke (1994), and Kestner (1996), in a quest to find risk measures that are relevant to investors and has been analyzed in different contexts in the academic literature (Grossman and Zhou, 1993; Alexander and Baptista, 2006; Schuhmacher and Eling, 2011; Kardaras et al., 2017; Roche, 2019).

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.

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.[3] 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.[4] 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.

Apart from their immediate application as risk measures, drawdown measures are important building blocks for performance ratios. Such ratios are commonly constructed for performance measurement purposes by dividing excess returns (over the risk-free rate) by some measure of risk (Caporin et al., 2014).[5] Using drawdown measures in the denominator, some of the

---

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

[4]Since Jensen (1968), a vast body of literature has studied manager skill using different empirical techniques (Grinblatt and Titman, 1989; Fama and French, 2010; Sénéchal, 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.

[5]This plug-in technique is inspired by the Sharpe ratio and may be theoretically sound for some risk measures in

resulting drawdown-based performance ratios are already known from the literature and have received names of their own; for example, *Calmar ratio* when the maximum drawdown is in the denominator, *Pain ratio* when the average drawdown is in the denominator, and *Ulcer ratio* or *Martin ratio* when the average squared drawdown is in the denominator (Cogneau and Hübner, 2009; Schuhmacher and Eling, 2011). 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 literature.[6] 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. 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.

2.2 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

---

some circumstances, like the Sharpe ratio under normally distributed returns (Sharpe, 1966, 1975, 1994). However, there is so far no theoretical justification to apply this technique to drawdown measures.

[6]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.

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 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:

$$\text{wDD} \coloneqq \sum_{i=1}^{N} \omega_i D_i, \qquad 0 \le \omega_i \le 1, \ \sum_{i=1}^{N} \omega_i = 1,$$

where $D_i \coloneqq \frac{M_i - S_i}{M_i}$ and $M_i \coloneqq \max_{t=0,\ldots,i} S_t$. 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 $K = \sum_{j=1}^{N} \frac{1}{N} D_j$. Such rescaling

---

[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. (2017b) 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.

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. 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 weight zero.

Figure 2.1 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 ADD² usually attach non-zero weights to most elements of the drawdown graph, MDD and 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,

---

[10]In the literature, for example in Caporin et al. (2014), the ADD² 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.
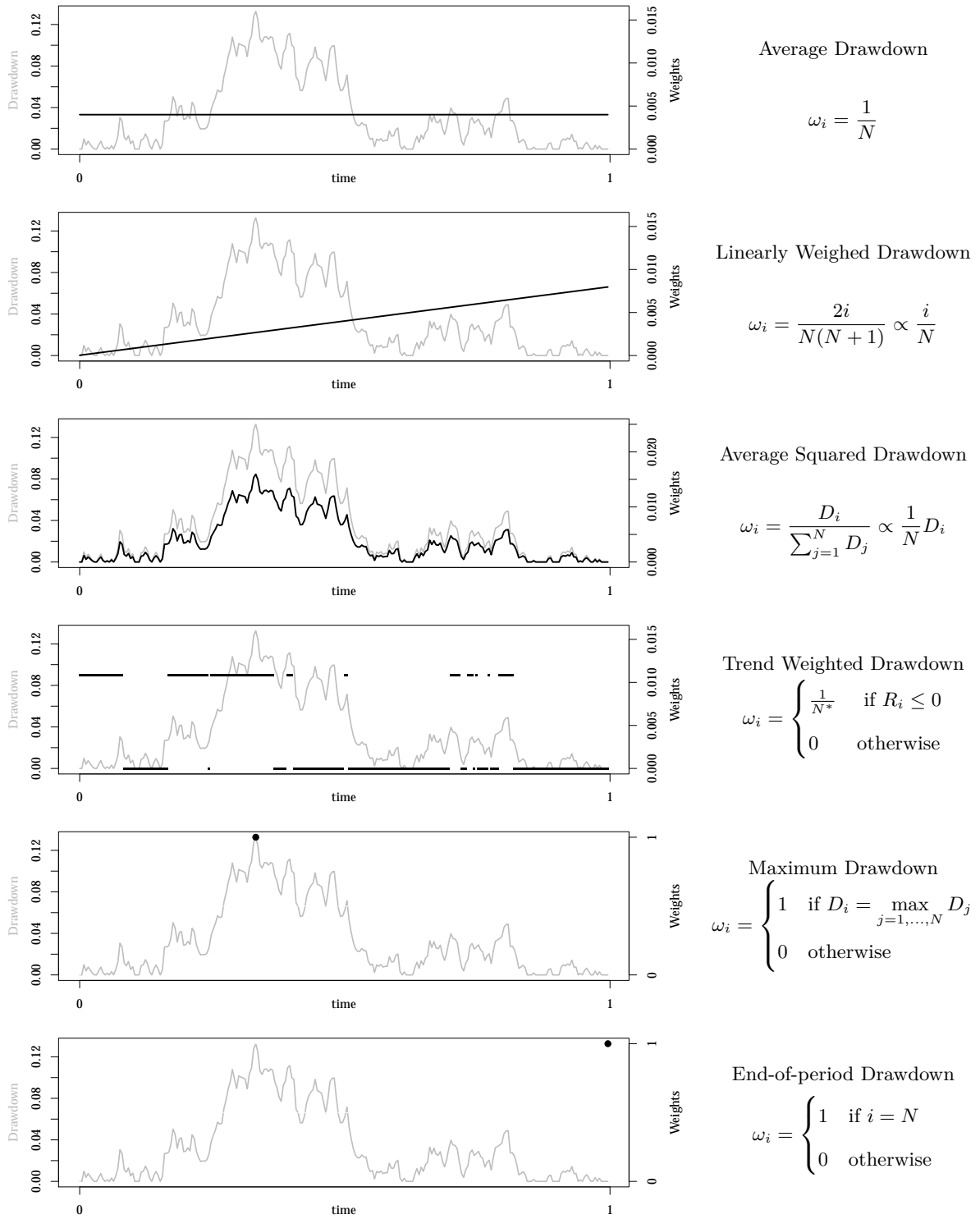
Figure 2.1: Illustration of Different Drawdown Measures within the wDD Framework. It shows the weights ω_i needed to obtain specific drawdown measures (i.e., ADD, lwDD, ADD^2, twDD, MDD, eopDD) within the wDD framework. It uses the same simulated drawdown graph with $N = 250$ for all drawdown measures. This drawdown graph is depicted by the grey lines. The weight functions are shown as black lines or black dots.

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 2.1 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.

2.3 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 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

---

[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.

[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).

of one year are obtained from Datastream. In total, our investment universe comprises 3,489 stocks from 26 countries and 11 sectors.[14]

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.[15] 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 $\{0\%, 0.1\% \dots, 2\%\}$, 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 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. Since the managers may select any stock from the index at the time of portfolio construction, the procedure is free from survivorship bias.

At the end of the month, the managers remove all stocks from the portfolio that have exited the

---

[14]This universe includes all stocks present in the MSCI World index at least once between December 1999 and April 2019. At the end of the horizon, the index comprises 1641 stocks from 23 countries and 11 sectors.

[15]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.

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.[16] 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.[17] In our simulation study, we vary δ between 50% and 60%, where 50% corresponds to the purely random case and 60% 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 2.1). Because of its effect on mean returns, higher manager skill also reduces the expected shortfall and the value-at-risk.

---

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

[17]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.

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 2.1: Summary Statistics of Portfolio Returns for Different Skill Levels. 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.

2.4 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.[18] We conduct pairwise comparisons for all measures mentioned in Figure 2.1, and include the standard deviation and the expected shortfall for comparison.[19] The objects to be ranked are the portfolios of 1,000 fictitious portfolio managers as described above. Rank correlations are obtained using Kendall's τ.[20] For each of the 210 (overlapping) one-year periods[21] in our data period, we compute the

---

[18]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)).

[19]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.

[20]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.

[21]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

rank correlation between two risk measures and report the time-averaged rank correlations in the lower triangle of Table 2.2. 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 2.2 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.

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 2.2: Rank Correlations Between ADD, lwDD, ADD^2, twDD, MDD, eopDD, Expected Shortfall (ES) and Standard Deviation (SD). 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.

The results in Table 2.2 have a straightforward interpretation. When someone compares the 1,000 unskillful portfolio managers using the ADD and someone compares them using the lwDD

---

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.

during a one year period, both rankings would (on average) be correlated by 0.840 (see Table 2.2, 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 2.2 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 2.1, 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 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.

Finally, we investigate whether similarities in portfolio rankings change if we rank portfolios according to drawdown-based performance ratios instead of drawdown measures. Compared to the rank correlations obtained from drawdown measures, as shown in Table 2.2, drawdown-based performance ratios lead to substantially higher values. The general increase in rank correlations is likely due to the common numerator of all performance ratios, the excess return. Nonetheless, all patterns present in Table 2.2 remain intact and the intuitions of the wDD framework remain valid. For example, the ratio with ADD in the denominator is still closest related to the ADD^2 ratio (correlation of about 0.94) followed by the lwDD, twDD and MDD ratios. The ratio with

eopDD in the denominator exhibits the lowest rank correlations of about 0.47.[22] In summary, drawdown-based performance ratios rank managers more similar than the drawdown measures themselves, but still display important differences that are in line with the wDD framework.

2.5 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. We evaluate the skill detection in each of the 210 overlapping one year periods. For each evaluation period, 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 negative correlation. 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 2.2 shows how accurately different drawdown measures discriminate between managers with more and less skill. To arrive at the boxplots in Figure 2.2, we use a one year rolling window with monthly steps, 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,

––––––––––––––––––––––––––––

[22]The full results are available from the authors upon request.
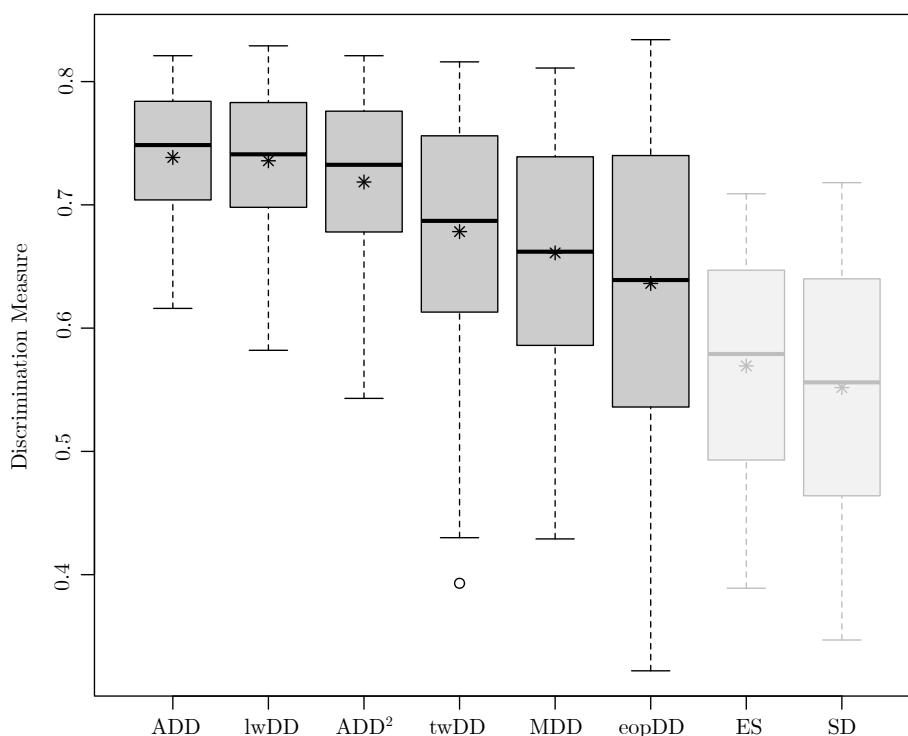
Figure 2.2: Discrimination Between Skillful and Unskillful Portfolio Managers Using Drawdown Measures of Risk. Each boxplot illustrates how the corresponding risk measure discriminates between skillful and unskillful managers. Each month for approximately 20 years of data, we observe the performance 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).

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 2.2. However, variation in this difference shows that the relative performance of the various drawdown measures remains unchanged. For reference purposes, we also include the

expected shortfall and the standard deviation[23], which only have little power to discriminate between skillful and unskillful managers.

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).

Given these results, it is a straightforward idea to exploit the drawdown measures' ability to detect skill for the improvement of performance ratios. Therefore, we investigate whether drawdown-based performance ratios, which we henceforth refer to with an additional subscript r, are particularly capable of detecting management skill. We report the corresponding proportions of correctly classified portfolio managers in Figure 2.3, which reveals several interesting findings. First, compared to the corresponding results for the drawdown measures in Figure 2.2, 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 eopDD$_{\mathbf{r}}$, which performs considerably worse. This is likely due to the following property of 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 standard deviation. Fourth, somewhat surprisingly, ADD$_{\mathbf{r}}$, lwDD$_{\mathbf{r}}$, twDD$_{\mathbf{r}}$, ADD$_{\mathbf{r}}^2$, and eopDD$_{\mathbf{r}}$ 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

---

[23]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).
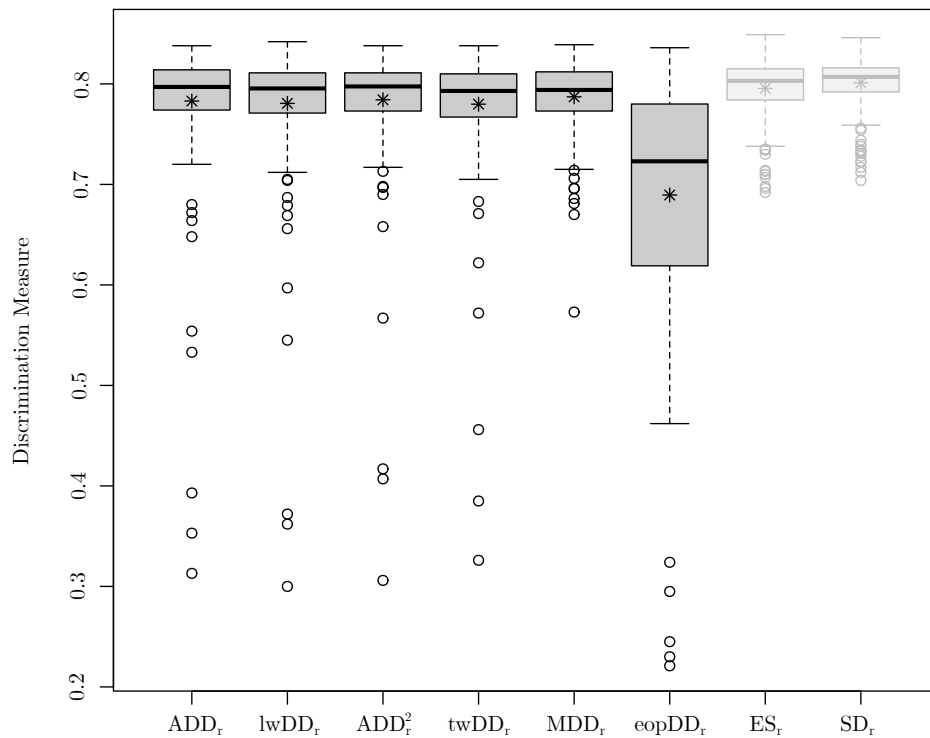
Figure 2.3: Discrimination Between Skillful and Unskillful Portfolio Managers Using Drawdown-based Performance Ratios. Each boxplot illustrates how the corresponding performance ratio discriminates between skillful and unskillful managers. Each month for approximately 20 years of data, we observe the 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).

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, 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.[24] 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

---

[24]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, Akeda (2003) and McLeod and van Vuuren (2004), and against, for example, Ferruz Agudo and Sarto Marzal (2004), Israelsen (2005) and Scholz (2007).

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.

2.6 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.

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 in line with the intuition from the wDD framework. For example, the most prominent drawdown measure, MDD, incorporates just the maximum of the drawdown graph and, hence, disregards plenty of drawdown-related information. Compared to the other drawdown measures in our study, it produces rankings that are most similar to those of the standard deviation, and it is not particularly strong at detecting skill. One of the simplest but rarely used measures – the average drawdown – shows the best results and we like to encourage investors, portfolio managers and academics to pay more attention to it.

When it comes to the application of drawdown in performance measurement, drawdown-based ratios detect skill 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.

One potential way of dealing with these problems is to apply additive combinations of risk and return measures instead of ratios.

Drawdown measures of risk could also be applied in portfolio optimization, for example, by maximizing drawdown-based performance ratios. One interesting question regarding this application is how the use of alternative drawdown measures affects the composition of optimal portfolios. Such optimization is challenging, however, because no formulas for the drawdown of a portfolio exist and numerical optimization may be infeasible for portfolios with a sizable number of assets. Improvements on this problem are an important issue for future research.

The simulation framework of our study captures signals relevant to the investment process in a controlled way. In this paper we consider a pure stock selection signal only. Future research could investigate the impact of allocation signals that can be implemented via additional segment constraints. One could also calculate the required hit ratio for a market timing signal to be profitably integrated into the investment process. Another idea is to investigate the minimal accuracy for applying a stock selection filter as a means to add value. Such analysis could provide guidance on the required signal accuracy for each individual step in the investment process.

B 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 $\text{risk}_i \geq 0$. We express $\text{ret}_2 = \text{ret}_1 \cdot \alpha$ and $\text{risk}_2 = \text{risk}_1 \cdot \beta$. Since we want to characterize when the skillful manager obtains the worse ratio despite having higher returns and lower risk, i.e., when $\text{ratio}_1 < \text{ratio}_2$, $\text{ret}_1 > \text{ret}_2$ and $\text{risk}_1 < \text{risk}_2$, we always have $\beta > 1$. We distinguish between three distinct cases:

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

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

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 $\text{ret}_2 > 0$, both returns are positive and $\alpha < 1$. Analogously,

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

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

(iii) If $\text{ret}_1 > 0 > \text{ret}_2$, 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.

3 Drawdown Persistence and a Convenient Shortcut to Predicting Mutual Fund Drawdown

Abstract

Drawdown risk measures allow investors to add an important path-dependent perspective to their assessment of risk. Computing drawdown measures to evaluate past performance is straightforward. However, whether drawdown measures can inform investment decisions critically depends on their persistence, i.e., whether funds' historical drawdown is indicative of their future drawdown. We find strong evidence of such persistence in an extensive sample of mutual funds for a wide variety of fund types and time horizons. While past drawdown is indicative of a fund's future drawdown, past volatility is even more informative, thus providing a convenient shortcut to predicting drawdown. Both further empirical analysis and simulations explain this surprising finding. They reveal that drawdown persistence is driven primarily by volatility persistence but is particularly sensitive to persistence in returns. Additionally, the results allow us to specify in which circumstances relying on either drawdown persistence or the shortcut would be more advantageous.

3.1 Introduction

Drawdown measures are among the most frequently applied risk measures in the asset management industry and are used, for example, to evaluate fund managers, to quote performance, or to aid fund allocation or redemption decisions (Landriault et al., 2015; van Hemert et al., 2020). Scores of funds use favorable drawdown statistics to attract institutional and private investors alike (Lhabitant, 2004). Commodity trading advisors are even mandated by the Commodity Futures Trading Commission to disclose drawdown statistics. From an investor's perspective, trying to manage future drawdown makes sense because drawdown measures excel at capturing what investors truly perceive as 'risk' (Harris et al., 2015).

On the lookout for funds with attractive drawdown properties, one's natural first guess may be to consider funds with an appealing drawdown track record. However, such statistics would be of little use if drawdown were not persistent, akin to placing lottery bets relying on the ticket numbers from last week's draw. So far, no evidence of drawdown persistence exists in the literature. Thus, a fund portrayed as having low historical drawdown may very well not be less risky in the future than its peers with high historical drawdown. Questions of persistence have been addressed for other risk measures in the literature (e.g., in Casarin et al. (2005)) – but not for drawdown. As Goldberg and Mahmoud (2017) note, despite their widespread use in practice, drawdown measures are far less developed in the literature than other risk measures. Therefore, we address this pressing question because using historical drawdown to guide investment decisions (e.g., when it is printed in investment brochures) is not logically sound unless drawdown is persistent.

From an ever-growing set of available drawdown measures, we focus on two of the most widely propagated measures, the maximum drawdown and the average drawdown. To assess if drawdowns of mutual funds are persistent, we use an extensive amount of data on U.S. mutual fund returns from 1990 to 2019 for more than 7,000 distinct funds. The two independent methods we use to assess relative persistence – one based on rank correlations, the other based on quantile portfolios – are among the most common approaches of quantifying persistence in the literature (e.g., Harri and Brorsen (2004), Busse et al. (2010)). Our principal finding is that drawdown measures are highly persistent. On average, funds with low historical drawdown have relatively low future drawdown. We document robust persistence for different fund categories, different time horizons, and different subperiods. Compared to the maximum drawdown, the average drawdown is equally as persistent, whereas in both cases the effect is highly economically significant.

When computing drawdown measures, it is imperative to specify the time window during which they are measured since it is not yet understood how results for short time horizons translate to long time horizons and vice versa. We compute all results for one year, for three years, and for three months, obtaining similar persistence for all. As deliberate risk-taking is often desired, we want to stress that these results can be applied not only to identify low-risk funds but also to steer future drawdown more generally across the whole drawdown spectrum.

As mutual fund drawdown is persistent, past drawdown can be utilized to predict future drawdown. However, we identify a ubiquitously available measure that is an even better predictor of future drawdown: past return volatility. With remarkable robustness, past standard deviation of returns predicts future maximum and average drawdown. Hence, anyone seeking to manage future drawdown should look first to the standard deviation, as it may provide a potent and convenient shortcut to predicting future fund drawdown.

To explain the – at first glance surprising – finding of the shortcut as the more accurate indicator, we analyze the data in more detail to extract the key drivers of drawdown persistence. Volatility persistence appears to drive drawdown persistence, but return persistence, if present, is an even stronger driver. In the absence of significant return persistence, past volatility is the superior drawdown indicator. These findings are underscored and extended in a simulation study with a stochastic mean stochastic variance model in which return and volatility persistence are explicitly included via long-memory processes. Return persistence emerges as the dominant factor that determines whether historical drawdown or standard deviation is the superior predictor.

By exploiting drawdown persistence or the volatility shortcut, future drawdown risk can be reduced by favoring funds with low historical drawdown or standard deviation. What are the implications of such a strategy? For example, are future returns different for funds with relatively low historical drawdown and funds with high drawdown? Empirically, the mutual fund returns in our data set suggest no such relationship as they remain largely unchanged. Hence, on average, choosing a fund with low drawdown or standard deviation does not hurt future returns. Similarly, the sorted funds do not differ significantly in terms of alpha with respect to the Fama-French 5-factor model. Only the market beta is lower for low-risk funds than for their high-risk counterparts. Thus, the simple approach to curb drawdown risk by considering historical drawdown or using the shortcut sacrifices neither returns nor alpha.

As the implications of our results for investment practice may seem conflicting, we summarize the key take-aways as follows. On the one hand, there is first-time evidence that considering

drawdown information in the investment process is useful in managing future drawdown properties. Therefore, a valid conclusion would be that, on average, if you pick a low drawdown mutual fund today, you can expect relatively low drawdown in the future. As the evidence is robust, our results support picking mutual funds based on historical drawdown. On the other hand, as identifying low drawdown funds via the volatility shortcut is more effective, the drawdown information seems redundant. Indeed, when return persistence is low, using the shortcut is a viable option; however, historical drawdown is the more accurate predictor if substantial return persistence can be expected. Therefore, whether to use the shortcut depends on the application. In any case, readily accessible information is available to improve the choice of mutual funds when future drawdown is of interest.

This paper is structured as follows. Section 3.2 explains how this paper contributes to the literature regarding drawdown measures and persistence. Section 3.3 provides details about the mutual fund data set, and Section 3.4 focuses on the empirical investigation of drawdown persistence. Section 3.5 introduces the standard deviation as a valid shortcut to predicting mutual fund returns, and Section 3.6 explains two approaches to understanding the drivers of persistence. The interplay between return, volatility, and drawdown persistence is explored in Section 3.6 first within the mutual fund data and then in a simulation study with fractionally integrated returns. Before the paper concludes, Section 3.7 addresses how managing drawdown affects average returns and alphas.

3.2 Literature Review

The concept of drawdown is far from new. In the 1980s, Garcia and Gould (1987) note that in their experience, many investors put more emphasis on maximum drawdown as a risk metric than on volatility. Many studies have since contributed to a better understanding of maximum drawdown. For example, Douady et al. (2000) and Magdon-Ismail et al. (2004) derive the distribution and expected value of the maximum drawdown for a (drifted) Brownian motion; Camara Leal and de Melo Mendes (2005) compare the maximum drawdown of several index time series to the maximum drawdown of a fitted Pareto model; Casati and Tabachnik (2013) compare empirical distributions of maximum drawdown to distributions simulated with skewness and excess kurtosis; Cheridito et al. (2012) derive the distribution of maximum drawdown for stopped processes of class sigma; Kim (2018) compare portfolio sorts based on maximum drawdown, value-at-risk, and volatility; and van Hemert et al. (2020) compute comparative statics to address

how changes to return, volatility, length of time horizon, and autocorrelation affect the maximum drawdown.

Although the maximum drawdown is arguably the most prominent drawdown measure, numerous other drawdown measures have been introduced. Most notably, Chekhlov et al. (2005) propose the average drawdown as part of a class of drawdown measures called conditional drawdown (CDD), which also includes the maximum drawdown.[1] In contrast to the maximum drawdown, which captures only the largest drawdown, the average drawdown is computed as the average of all drawdowns. Similarly, Martin and McCann (1989) mention a measure where all drawdowns are first squared and then averaged. Some drawdown measures, like the average continuous drawdown, are used predominantly in the denominator of drawdown-based performance ratios, cf. Schuhmacher and Eling (2011). Bradford and Siliski (2016) propose the active drawdown measure which is computed relative to a benchmark index. In addition to measures based on the intensity of drawdowns, measures related to the duration of drawdowns have also been discussed, for example, in Mahmoud (2017). Goldberg and Mahmoud (2017) introduce the conditional expected drawdown (CED), which is useful as an ex-ante risk concept but not applicable to ex-post evaluation because it requires the distribution of maximum drawdowns and cannot be computed for a single sample path. A modification of CED with demeaned returns is proposed in Molyboga and L'Ahelec (2016). The end-of-period drawdown measure, which emphasizes aspects of regret at the end of the evaluation period, is introduced in Möller (2018). Korn et al. (2019) propose two new measures, the trend weighted drawdown and the linearly weighted drawdown, as part of a comprehensive weighted drawdown framework, which includes many previous drawdown measures.

Apart from the development of drawdown measures, several other strands of drawdown literature have developed. For example, Grossman and Zhou (1993) incorporate a drawdown constraint into a continuous-time investment problem in which a specified drawdown must not be exceeded at any time. This application has attracted significant attention in the literature, including by Cvitanic and Karatzas (1995), Alexander and Baptista (2006), Elie and Touzi (2008), Sekine (2013), Yao et al. (2013), Cherny and Obłój (2013), Rieder and Wittlinger (2014), Angoshtari et al. (2016), Kardaras et al. (2017), and Roche (2019), whereas the drawdown constraint has been generalized, other constraints have been added, and results have been extended to

---

[1]In general, the CDD class includes drawdown measures where the drawdown with respect to the running maximum is continuously assessed, and the worst $(1 - \alpha) \cdot 100\%$ of these drawdowns are averaged for α between zero and one.

different underlying processes and portfolios. A different strand of largely mathematical literature has assessed stochastic properties of the drawdown process, see Hadjiliadis and Vecer (2006), Mijatović and Pistorius (2012), Landriault et al. (2017b), and Bai and Liu (2019). Another area of ongoing debate is whether different drawdown-based performance ratios lead to different rankings of investments, see Eling and Schuhmacher (2007), Eling (2008), Caporin and Lisi (2011), Haas Ornelas et al. (2012), Auer and Schuhmacher (2013), Auer (2015), and Korn et al. (2019).

Additionally, other ideas regarding drawdown have been pursued. For example, Vecer (2006) and Vecer (2007) investigate the relation between drawdown and option pricing, Heidorn et al. (2009) use the maximum drawdown to analyze risk properties of funds of hedge funds, Gilli and Schumann (2009) employ drawdown measures among other alternative risk measures in portfolio optimization, and Pospisil and Vecer (2010) define and analyze drawdown Greeks. Moreover, Zabarankin et al. (2014) develop drawdown-β and drawdown-α with respect to a drawdown CAPM, Palmowski and Tumilewicz (2017) price drawdown-type insurance contracts, and Challet (2017) use the drawdown duration to construct an estimator for the Sharpe ratio. Although drawdown measures are included in many surveys of risk or performance measures (e.g., Bacon (2008), Caporin and Lisi (2011), and Caporin et al. (2014)), they remain not as well studied as other risk measures, such as value-at-risk or expected shortfall.

Let us turn to the literature on persistence. Because of its fairly broad literal meaning, i.e. that a phenomenon continues to exist for a prolonged period, questions of persistence have been addressed in vastly different fields of finance and economics. These include the persistence of inflation (Pivetta and Reis, 2007), the persistence of firm capital structure (Lemmon et al., 2008), the persistence of bank profits (Goddard et al., 2011), the persistence of executive compensation (Cheng et al., 2015), and the persistence of earnings, cash flows, and accruals (Hui et al., 2016). However, the arguably most prominent strand of the persistence literature is concerned with volatility persistence, also known as 'volatility clustering'.[2] Its fundamental observation dates back at least to Mandelbrot (1963) who notes that large price movements typically follow previous large movements of either sign, and small changes typically follow previous small changes of either sign. More rigorously, Ding et al. (1993) observe that autocorrelations of absolute and squared returns – or, more generally, $|r_t|^d$ – are positive for various exponents d and even for long lags. The study of volatility clustering has benefited immensely from the development of (G)ARCH

---

[2]We use the terms volatility persistence, volatility clustering, and persistence in standard deviation interchangeably.

models by Engle (1982) and Bollerslev (1986), which explicitly model persistence in volatility. Utilizing these models, volatility persistence has been documented with high-frequency as well as low-frequency data ranging from intraday to monthly returns (Chan et al., 1991; Jacobsen and Dannenburg, 2003).

While the results on volatility persistence are fairly unanimous, return persistence – or, more generally, performance persistence – is surrounded by much more ambiguity. Several studies in the early 1990s support claims of persistence. For example, Jegadeesh and Titman (1993) demonstrate persistence in single stocks by showing that strategies that buy past winners and sell past losers generate significant positive returns. Assessing mutual funds,[3] Grinblatt and Titman (1992), Hendricks et al. (1993), Brown and Goetzmann (1995), and Elton et al. (1996) find evidence for performance persistence in terms of returns and alpha against different portfolio benchmarks. To measure persistence, they consider regression results, portfolio sorts, contingency tables, and rank correlations, respectively. In a seminal paper, Carhart (1997) finds that persistence diminishes if momentum is taken into account. First, he replicates that portfolios sorted on past return differ substantially in return and CAPM alpha in the following year. Then, he demonstrates that these differences disappear when alpha is computed with respect to a four-factor model, which includes a factor-mimicking portfolio for one-year return momentum. Incorporating this finding by measuring performance as alpha with respect to Carhart's four-factor model, the subsequent literature contains mixed results. Bollen and Busse (2005) find short-term persistence in daily returns that disappears for longer horizons. Cohen et al. (2005) report persistence in momentum-adjusted returns and report even stronger persistence in a new performance measure that includes holdings information relative to other funds. Kosowski et al. (2006) find some evidence of persistence using a bootstrap approach, whereas Fama and French (2010) use a slightly different bootstrap procedure and find hardly any evidence of persistence. Huij and Verbeek (2007) find some persistence, especially for small-cap or growth funds, when they improve the sorting into portfolios with an empirical Bayes approach. Barras et al. (2010) argue that previous approaches do not distinguish between superior performance because of skill or because of luck; they try to build portfolios with funds that truly have skill but do not find substantial

---

[3]The majority of the performance persistence literature studies mutual funds. Analyzing hedge funds, Ammann et al. (2013) report significant performance persistence, and Eling (2009) find that persistence critically depends on the type of hedge fund. Analyzing portfolios managed by institutional investment management firms, Busse et al. (2010) find that modest persistence is present in three-factor alphas but disappears after controlling for momentum.

outperformance. Berk and van Binsbergen (2015) find persistence in mutual fund performance when combining the four-factor alpha with assets under management to obtain a 'value-added' performance measure. El Ghoul and Karoui (2017) compare funds with high and low corporate social responsibility (CSR) scores and observe that high-CSR funds exhibit stronger performance persistence than low-CSR funds. Harvey and Liu (2018) use a panel regression framework to estimate fund alphas and find some evidence of persistence. Overall, performance persistence is a topic of ongoing debate, and at least some persistence can be observed.

The persistence of higher-order moments has been investigated as well. Ex-post stock returns exhibit positive skewness, which has been found to persist. For example, Singleton and Wingender (1986) observe the skewness of monthly stock returns to be weakly persistent over consecutive five-year periods by computing rank correlations and transition frequencies. Using the same data, Muralidhar (1993) conducts a bootstrap test that considers the sampling distribution of the sample skewness and concludes that skewness is strongly persistent. Defusco et al. (1996) and Nath (1996) support these results by extending the data coverage and adding a new bootstrap test based on the sampling distribution of the difference in skewness. While the skewness persistence of individual stocks does not automatically carry over to portfolios, Sun and Yan (2003) find that mean-variance-skewness efficient portfolios exhibit skewness persistence. Jondeau and Rockinger (2003) document persistence for the conditional skewness and kurtosis, which they compute with parameter estimates of a GARCH model with generalized t-distributed residuals. For time series of stock indices and exchange rates, they find that both conditional skewness and kurtosis are persistent, but skewness persistence exceeds kurtosis persistence. Ergün (2011) analyzes the same question with robust skewness and kurtosis measures and obtains mixed results.

We pick up both strands of persistence and drawdown literature and examine the persistence of drawdown measures. From the persistence angle, expanding the set of analyzed quantities to the increasingly important class of drawdown measures appears worthwhile. From the drawdown angle, addressing persistence is crucial to substantiate the use of drawdown measures in investment practice.

3.3 Mutual Fund Data

In 2018, every third American – or 43.9% of American households – owned shares of mutual funds, according to the Investment Company Institute (2019). With assets under management of approximately 10 trillion U.S. dollars (USD), equity mutual funds constitute a pivotal investment vehicle for private and institutional investors alike. Our sample includes all mutual funds

domiciled in the U.S. between January 2, 1990 and September 25, 2019 from the Morningstar database (used, e.g., by Berk and van Binsbergen (2015) and Ma et al. (2019)). We focus on equity funds and omit money market funds, bond funds, and funds that trade commodities. The sample includes defunct, alive, and new funds, thus avoiding survivorship bias. Many fund companies sell funds with multiple share classes. These may differ in fee structure, minimum investments, or distribution channel, but not in their investment portfolio (Nanda et al., 2009). As their returns are almost perfectly correlated, including all share classes of a fund would add little information but would artificially increase the sample size and statistical significance. Different share classes of the same fund are identified using Morningstar's fund ID.[4] When multiple share classes are present, we only retain the total return index for the primary share class.

To group funds into categories, we use the Morningstar category, which is a quantitative and qualitative assessment of a fund's three-year primary investment focus. On a broader scale, several Morningstar categories map to a Morningstar category group. This allows us to conduct analyses on three levels of detail: the full sample, the Morningstar category groups, and single Morningstar categories. The final sample comprises funds from 40 Morningstar categories and three Morningstar category groups: U.S. equity (4,636 funds), international equity (2,064), and sector equity (781). Morningstar categories sometimes change in time, but changes from one Morningstar category group to another are rare. We omit the 'Morningstar' attribute when writing about categories from here onward. To ensure that only investable funds are included, funds are not considered until they reach a size of $5 million in assets under management (AUM) as in Berk and van Binsbergen (2015).[5] We also disregard portions of the path where the data is not yet available at the daily frequency.

The Morningstar database lists 33,253 equity funds for our selected period. After consolidating multiple share classes, the sample includes 9,394 funds, 7,263 of which have data available at the daily frequency and sufficient assets under management. Of these funds, 3,585 are alive, and 3,678 are defunct at the end of the data period. The evolution of the number of funds in the data

---

[4]While Berk and van Binsbergen (2015) use the last letters of the fund names to identify subclasses, we use the fund ID supplied by Morningstar. Comparing both approaches by a rough spot-check, we verify that the fund ID finds all subclasses identified by the approach of Berk and van Binsbergen (2015) but not vice versa.

[5]Ma et al. (2019) exclude more funds by applying a threshold of $15 million. As a robustness check, we conduct our analysis again for the even more conservative threshold of $50 million AUM. All results change only marginally, and all conclusions remain valid.

set over time and their aggregated assets under management are displayed in Figure 3.1a. While the number of funds has risen steadily except for a short period after the Great Recession, assets under management have been more volatile, finally exceeding $10 trillion. For each fund, we obtain the total return index as a measure of the fund's return to an investor when all dividends and distributions are reinvested. The mean total return is 6.5% p.a. at a standard deviation of 20.2% with slight left-skewness and excess kurtosis, see Figure 3.1b.

	All	U.S. Equity	Int. Equity	Sector Equity
Mean	0.065	0.072	0.051	0.056
Std Dev.	0.202	0.199	0.185	0.244
Skewness	−1.178	−1.252	−0.919	−1.206
Kurtosis	8.722	10.276	6.096	7.301
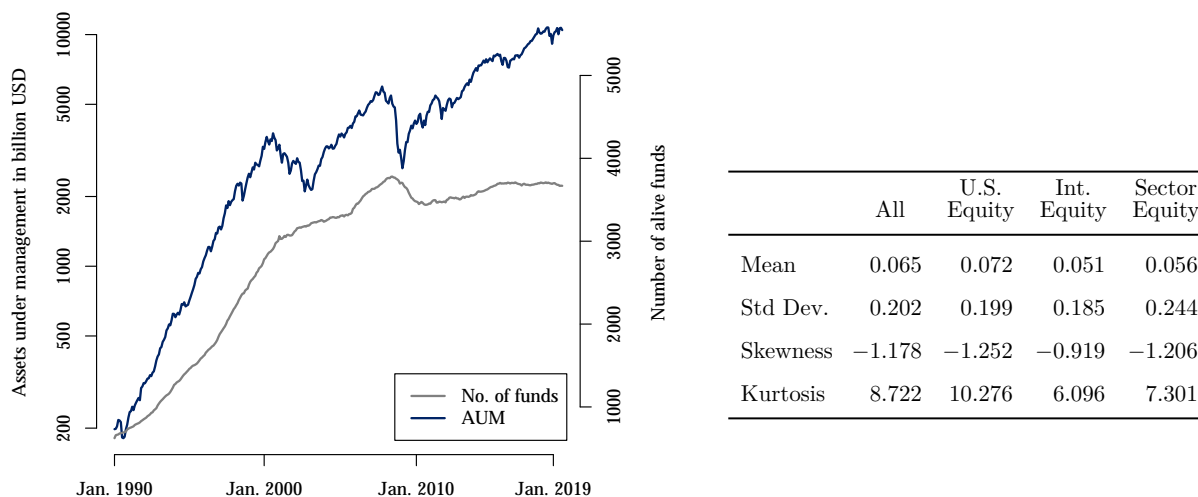
(a) Extent of the data set in terms of assets under management (AUM) in billion USD (left scale, log scaled) and the number of alive funds (right scale).

(b) Moment properties of the total returns. Mean and standard deviation are annualized estimates from daily log returns. Skewness and kurtosis are based on annual log returns.

Figure 3.1: Summary statistics of the mutual fund data set.

3.4 Drawdown Persistence

Imagine facing a choice between two mutual funds, both of which provide their drawdown track record. Does incorporating this information about past drawdown improve the investment decision? The answer to this question critically depends on drawdown persistence. If drawdown is not persistent, factoring in this information may do more harm than good. Conversely, if past drawdown *is* persistent, it contains potentially valuable information about future drawdown, and investors would be imprudent to dismiss this information. For example, picking the fund with lower drawdown – given all else is equal – may reduce drawdown risk in the future. More generally, if the relationship between past and future drawdown is roughly proportional, past drawdown may be used to steer future drawdown toward the desired risk–return characteristic. However, it is all in vain without drawdown persistence, which is why we address it thoroughly in this section. As a prerequisite, we first note how to define drawdown measures adequately in the following.

3.4.1 Definition of Drawdown Measures

In general, drawdown measures capture losses from previous peaks. As its name suggests, the maximum drawdown (MDD) selects the largest loss from a previous peak. It is highly intuitive, as it quantifies the worst possible market timing (i.e., buying and selling an asset at the worst combination of buying high and selling low). In the investment industry, this peak to subsequent trough measure has a firm foothold at least since the 1980s (Garcia and Gould, 1987). For observed daily market values $S_0, S_1, \ldots, S_N$ of an investment, denote the running maximum by $M_t := \max_{i=0,\ldots,t} S_i$. The MDD is then given as the maximum of all drawdowns $D_t := \frac{M_t - S_t}{M_t}$. For each point in time t, the drawdown D_t denotes the current percentage loss from the previous running maximum; its path is termed the 'drawdown graph,' and it contains comprehensive information regarding when, for how long, and how deeply an investment has experienced drawdowns. By picking the maximum, the MDD highlights a single element of the drawdown graph. Alternatively, the average drawdown (ADD) incorporates all elements of the drawdown graph equally, as it is defined as the average of all D_i. As it tracks by how much an investment has been in drawdown on average, it is usually much smaller than the MDD. Both measures belong to the conditional drawdown family introduced by Chekhlov et al. (2005), which contains risk measures defined as tail means of the highest $(1 - \alpha) \cdot 100\%$ of drawdowns. Within this family, the MDD results from letting the confidence parameter α tend to one, whereas the ADD results from setting α to zero. Not only within the conditional drawdown family but also in empirical asset rankings, MDD and ADD have been found to come from quite opposite ends of the spectrum of drawdown measures (Korn et al., 2019). Therefore, the persistence results for MDD and ADD should be fairly representative for drawdown measures in general. In summary, we apply the maximum drawdown and average drawdown defined by

$$\text{MDD} = \max_{t=1,\ldots,N} D_t \qquad \text{and} \qquad \text{ADD} = \frac{1}{N} \sum_{t=1}^{N} D_t.$$

By construction, drawdown measures are path-dependent downside risk measures, which are asymmetric and incorporate loss aversion relative to a time-varying benchmark. Of course, drawdown measures are always computed for time spans instead of static points. For example, we compute the MDD for two consecutive periods $[t-1, t]$ and $[t, t+1]$, which we abbreviate to $t-1$ and $t+1$, respectively. The maximum drawdown for the corresponding periods is denoted by MDD_{t-1} and MDD_{t+1}, respectively.

3.4.2 Quantifying Persistence

In order to quantify persistence, we apply two different techniques. The first, which is based on computing correlation coefficients, has the advantage of aggregating all information into a single intuitive number. The second technique, which relies on constructing quantile portfolios, allows for a more detailed look across risk levels and provides a grasp of the effect's order of magnitude. Let us describe both approaches in more detail.

If a property is persistent, objects with low historical values tend to have low future values, and high historical values typically coincide with high future values. That is, historical and future values are positively correlated. Therefore, our first measure of persistence is a correlation measure. Correlations are always framed for pairs of values. In our setup, we compare the drawdown of the same fund in two consecutive time intervals. For now, both time intervals are 12 months, but different time intervals are discussed in detail later. We follow a rolling window approach and shift two consecutive one-year windows through the time series in monthly steps and collect all such pairs. To obtain relative drawdown measures for each fund, we do not consider each fund in isolation, but we first compute drawdown measures for all funds available at a certain point in time, and then rank all funds based on historical drawdown. The relative rank, for example, third-highest drawdown of 100 funds, is given by the rank divided by the number of available funds, i.e., $\frac{3}{100} = 0.03$. Similarly, the relative rank of each fund is computed for the second period t+1. We collect all such pairs of relative ranks for all funds and all points in time, and compute the Pearson correlation coefficient. As usual, positive values (up to 1) indicate a positive correlation, while negative values (up to -1) indicate anti-correlation. We denote this correlation-based persistence measure by $\mathcal{P}_{\text{cor}}$. It is known as Spearman's rank correlation measure and used to measure persistence, for example, in Elton et al. (1996) and Harri and Brorsen (2004). This correlation measure also has an alternative interpretation in terms of a regression slope. Once all pairs of relative ranks are computed, the time t+1 results can be regressed on the time t−1 results. The slope of the linear regression would indicate if the time t−1 information was related to the time t+1 outcome. The correlation measure $\mathcal{P}_{\text{cor}}$ and the regression slope are identical for a straightforward reason.[6] Therefore, $\mathcal{P}_{\text{cor}}$ can just as well be

---

[6] Let (x_i, y_i) denote the pair of relative ranks for the ith fund during t−1 and t+1. Let $\bar{x}$ and $\bar{y}$ be the corresponding averages across all funds. By definition, the correlation is given by $\sum_i (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}$ and the least squares estimate of the linear regression slope by $\sum_i (x_i - \bar{x})(y_i - \bar{y}) / \sum_i (x_i - \bar{x})^2$. Notice that the numerators coincide. Moreover, the denominators coincide if $\sum_i x_i - \bar{x} = \sum_i y_i - \bar{y}$. Since both x_i and y_i are relative ranks, i.e., $\frac{1}{N}, \frac{2}{N}, \ldots, 1$, their averages and both sums coincide as well.

interpreted as a regression slope when the relative ranks of time t+1 are regressed on the relative ranks of time t−1.

Although $\mathcal{P}_{\text{cor}}$ is an intuitive metric to quantify persistence, we apply a second metric that indicates not only the strength of persistence but also its economic scale. This measure is based on portfolio sorts. After observing the drawdown during period t−1, all mutual funds are sorted into deciles; then, the average future drawdown during time t+1 is observed within each decile. Drawdown persistence would lead to the funds in the lowest drawdown quantile portfolio $\mathcal{Q}_1$ also exhibiting the lowest drawdown during the subsequent period. Similarly, the funds sorted into the highest drawdown quantile portfolio $\mathcal{Q}_{10}$ during t−1 would experience high drawdown during period t+1. When analyzing quantile portfolios, one often quantifies effects by comparing the values for the two opposite quantile portfolios $\mathcal{Q}_1$ and $\mathcal{Q}_{10}$; the difference $10 - 1$ and the fraction $10/1$ are common examples. Since the values for ADD are typically much smaller than the values for MDD, the relative $10/1$ measure is a good choice to compare effect sizes for MDD and ADD. We call this persistence measure $\mathcal{P}_{10/1}$. For example, those 80 out of 800 funds in $\mathcal{Q}_1$ with the lowest drawdown during 2018 might experience an average MDD of 20% during 2019. When those 80 funds in the highest drawdown quantile $\mathcal{Q}_{10}$ on average experience an MDD of 25% during 2019, the persistence measure $\mathcal{P}_{10/1}$ would equal $\frac{25}{20} = 1.25$. Of course, the intermediate quantile portfolios two to nine contain information as well. If drawdown persistence is strong, all portfolios between the two extremes will express drawdowns in t+1 that are monotonically increasing from one quantile portfolio to the next. Hence, the analysis with quantile portfolios assesses if the persistence effect is only present in the extremes or across all drawdown levels. The quantile portfolio approach is a widely accepted method to assess persistence, which is used, for example, in Carhart (1997) and Berk and van Binsbergen (2015).

Note that $\mathcal{P}_{\text{cor}}$ and $\mathcal{P}_{10/1}$ are far from identical. $\mathcal{P}_{\text{cor}}$ is a purely relative persistence measure that would attest to perfect persistence if the highest drawdown fund in t−1 exhibits the highest drawdown in t+1, the second-highest drawdown fund in t–1 exhibits the second-highest drawdown in t+1, and so on. This persistence measure does not provide information regarding the order of magnitude of the effect.[7] Conversely, $\mathcal{P}_{10/1}$ provides an intuitive understanding of the economic effect. For example, a value of 1.5 for $\mathcal{P}_{10/1}$ indicates that the highest drawdown decile on average experiences 50% more drawdown during t+1 than the lowest drawdown decile. Therefore, we

---

[7]It is possible that the ranking during t+1 perfectly coincides with the ranking during t−1, but the differences in drawdown are arbitrarily small and not economically significant.

always report both persistence measures. Regarding their interpretation, higher values indicate stronger persistence. For $\mathcal{P}_{\text{cor}}$, a value of zero is neutral; for $\mathcal{P}_{10/1}$, a value of one indicates neither positive nor negative persistence.

To avoid any misunderstanding, we quickly also note how we do *not* measure persistence. Considering the time series of a mutual fund in isolation, positive autocorrelation of drawdowns may be interpreted as persistence. Similarly, a positive regression slope of future drawdown regressed on historical drawdown values may also be understood as persistence. To exhibit such an absolute kind of persistence, past drawdown would have to exhibit strong capabilities in predicting future returns, as those strongly affect the future drawdown. However, predicting returns is notoriously difficult, and there seems little justification that past drawdown is a promising way of predicting future performance. Instead, we focus on fund persistence relative to peers for several reasons. Such relative persistence is more appropriate than absolute persistence, as mutual funds are often evaluated in relative terms such as tracking error or alpha with respect to a benchmark portfolio because investors in mutual funds generally accept some market risk. Additionally, when choosing mutual funds, investors are more likely to compare drawdown figures for different funds and observe how they fared in comparison, than to judge the absolute numbers. From a different perspective, assessing relative performance is fairer to a mutual fund manager because while accurately predicting market movements is challenging, tilting portfolios to be rather stable or risky is achievable. Our approach also ensures that results for different market phases are comparable. Moreover, even for relative persistence, there are various methods for quantification. For example, Malkiel (1995), Brown and Goetzmann (1995), and Mateus et al. (2019) utilize contingency tables, where funds are sorted into winners (W) and losers (L) in two subsequent periods and the frequency of the four combinations (i.e., WW, WL, LW, LL) is observed. When we sort funds into winners and losers according to above-median or below-median MDD in each year from 2000 to 2017, the categories WW and LL comprise more funds than the categories WL and LW, which indicates persistence. χ^2 tests and odds ratio tests are significant at the 99% confidence level in 17 and 16 out of 18 years, respectively. By disregarding all information except being above or below the median, this approach is far less information efficient than using, for example, the rank correlation measure $\mathcal{P}_{\text{cor}}$. Therefore, we do not report further results on contingency tables but report $\mathcal{P}_{\text{cor}}$ instead.

Before turning to the results, let us discuss a few technical details about the computation of $\mathcal{P}_{\text{cor}}$ and $\mathcal{P}_{10/1}$. When we compute the drawdown measures for the period $t-1$, we first check how many funds have return data available for the entire $t-1$ period. We proceed only if more funds

than twice the number of quantile portfolios are available. For all such funds, the return data for the t+1 period is obtained next. If returns should become unavailable for a fund during period t+1, the last obtainable price in the time series is used. In the tables below, we report a single value for $\mathcal{P}_{cor}$ as well as for $\mathcal{P}_{10/1}$ considering specific sets of funds across several years using the rolling window approach described above. For $\mathcal{P}_{cor}$, we obtain a single number because we collect pairs of relative ranks simultaneously across funds and across time, pool all such pairs, and compute the correlation coefficient. To obtain $\mathcal{P}_{10/1}$, we first compute the drawdown for all funds in the quantiles $\mathcal{Q}_1$ and $\mathcal{Q}_{10}$ separately for each period, and then take the average over time. Therefore, the drawdown values for $\mathcal{Q}_1$ and $\mathcal{Q}_{10}$ both result from a time-series average of cross-sectional averages. We take the ratio $10/1$ in the end because, statistically, a ratio of averages is a far more robust estimator than an average of ratios.

3.4.3 Results on Persistence

We investigate persistence for the maximum drawdown and the average drawdown of all mutual funds in the data set. The most general results are displayed in Table 3.1. For the MDD of the full sample, both $\mathcal{P}_{cor}$ and $\mathcal{P}_{10/1}$ indicate strong positive persistence. The value for $\mathcal{P}_{cor}$ indicates that the MDD ranking during t−1 and the MDD ranking during t+1 are correlated by 0.436; the value 1.74 for $\mathcal{P}_{10/1}$ results from the MDD of quantile portfolio 10 exceeding the MDD of quantile portfolio 1 by 74%. On average, quantile portfolio 1 exhibits an MDD of 12.7%, while quantile portfolio 10 exhibits an MDD of 22.1%, which is highly economically significant. The MDD increases monotonically from portfolio 1 to portfolio 10, which is indicated by the checkmark (✓). Therefore, persistence is present throughout the sample and not only in the extremes. The results for the ADD are similar, with $\mathcal{P}_{cor}$ being 0.409 and $\mathcal{P}_{10/1}$ being 2.02. Again, the drawdown increases from quantile portfolio to quantile portfolio, but the drawdown itself is smaller (between 4.1% and 8.3%) because an average, not a maximum, is considered. Table 3.1 also illustrates the results for U.S. funds, international funds, and sector funds. While all results are close to the results for the full sample, international funds typically exhibit lower persistence than sector funds. The number of funds for each subsample is indicated in the last column of Table 3.1.[8]

---

[8]The number of funds for the entire sample differs slightly from the sum of the three category groups because a few funds change the category group throughout the sample period.

Category Group	Persistence		Quantile Portfolios										incr.	# funds
	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10		
MDD:														
All Equity	0.436	1.74	0.127	0.139	0.143	0.146	0.150	0.157	0.164	0.173	0.185	0.221	✓	7263
U.S. Equity	0.465	1.61	0.125	0.134	0.138	0.140	0.145	0.149	0.157	0.165	0.177	0.201	✓	4636
Int. Equity	0.393	1.58	0.139	0.151	0.157	0.159	0.163	0.165	0.170	0.181	0.195	0.219	✓	2064
Sector Equity	0.468	2.40	0.122	0.142	0.160	0.165	0.169	0.179	0.192	0.215	0.242	0.293	✓	781
ADD:														
All Equity	0.409	2.02	0.041	0.045	0.047	0.049	0.051	0.055	0.057	0.060	0.065	0.083	✓	7263
U.S. Equity	0.441	1.77	0.039	0.043	0.044	0.046	0.048	0.050	0.053	0.056	0.060	0.069	✓	4636
Int. Equity	0.369	1.75	0.048	0.052	0.055	0.056	0.058	0.060	0.063	0.068	0.074	0.084	✓	2064
Sector Equity	0.400	2.88	0.042	0.051	0.056	0.057	0.060	0.062	0.070	0.083	0.096	0.121	✓	781

Table 3.1: Drawdown persistence results for the full sample. The top half reports on MDD persistence, the bottom half on ADD persistence. Each row corresponds either to the full sample (all equity) or to a subsample based on category groups. The persistence measure $\mathcal{P}_{cor}$ is the rank correlation between MDD_{t-1} and MDD_{t+1}; the measure $\mathcal{P}_{10/1}$ is the ratio $10/1$ of the quantile portfolios 1 and 10. To obtain quantile portfolios, funds are sorted into decile portfolios based on MDD_{t-1}, and the average MDD_{t+1} for each portfolio is computed across funds and across time. A checkmark (✓) indicates whether MDD_{t+1} increases monotonically from quantile portfolio 1 to portfolio 10. All definitions apply to the ADD accordingly. The number of funds in each subsample is denoted by # funds.

According to Table 3.1, drawdown measures exhibit persistence for the whole sample of mutual funds as well as for each category group. Examining more detail, we assess for each single fund category within each category group whether MDD and ADD persist. Categories with fewer than 100 funds are omitted from the analysis because results based on decile portfolios with such few assets may become unreliable. The results in Table 3.2 demonstrate that MDD and ADD are highly persistent in every single fund category. Even remarkably similar funds – which, for example, all invest in small growth stocks – benefit from being sorted by historical drawdown when information about the future drawdown is desired. Are there differences in persistence between the categories? U.S. funds tend to be more persistent than international funds, which is in line with the aggregated results for the category groups in Table 3.1. In the extremes, persistence varies between 0.465 for small value funds and 0.238 for foreign large growth funds measured with $\mathcal{P}_{cor}$, or between 1.50 and 1.19 using $\mathcal{P}_{10/1}$. For some categories, drawdown does not increase strictly monotonically between decile portfolios, especially when the number of funds in a given category is small. Furthermore, the absolute level of drawdown across the decile portfolios differs reasonably between categories: Large value funds exhibit less drawdown across the board compared to small growth funds, which in turn exhibit less drawdown than sector funds focused on the technology sector. For each size category – large, mid-cap, and small –

drawdown is highest for growth funds, lowest for value funds, and in between for blended funds. The strong persistence results for each category hold both for MDD and ADD.

Drawdown measures do not easily scale with time. While the MDD for 24 months is undoubtedly no smaller than the MDD for 12 months, there is neither law nor reliable estimate nor heuristic regarding how much larger the 24-month MDD can be expected to be. Therefore, each investor has to choose a suitable evaluation period when applying drawdown measures. Although for many investors, an annual horizon may be adequate, other choices are equally valid. To make sure that the persistence results do not depend on the annual evaluation period, all previous results are also computed for $t+1$ periods of three months and three years, and the results are provided in Tables 3.10, 3.11, 3.12, 3.13, and 3.14 in the appendix. All conclusions regarding persistence remain valid for both alternative choices of computing the drawdown measures.

Is drawdown persistence an old or a new phenomenon? Considering almost 30 years of data, was drawdown persistent only in the 1990s and never again since, or has its persistence (re)emerged recently? Robustness over time is tested by splitting the sample into three subperiods: 1990–1999, 2000–2009, and 2010–2019, each from January to either the end of the year or the end of the data period.[9] According to Table 3.3, strong persistence is present during all three subperiods. It is slightly stronger between 1990 and 1999, but there is no trend in general. If anything, persistence has been rising lately according to ADD. While the two full-grown financial crises between 2000 and 2009 cause higher absolute levels of drawdown in the second subperiod, drawdown persistence remains largely unaffected. Hence, drawdown persistence persists over time.

---

[9]Persistence is strongly present in all subperiods irrespective of the sample being split into three, four, or five subperiods.

Panel A: Persistence of MDD

Category	Persistence		Quantile Portfolios										incr.	# funds
	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10		
Large Value	0.370	1.33	0.113	0.123	0.127	0.130	0.132	0.132	0.135	0.138	0.140	0.150	✓	956
Large Blend	0.357	1.31	0.117	0.127	0.130	0.132	0.133	0.133	0.135	0.139	0.143	0.153	✓	1641
Large Growth	0.391	1.38	0.134	0.142	0.146	0.149	0.153	0.157	0.160	0.165	0.172	0.185	✓	1144
Mid-Cap Value	0.409	1.39	0.118	0.129	0.133	0.137	0.139	0.141	0.144	0.146	0.151	0.164	✓	456
Mid-Cap Blend	0.412	1.45	0.126	0.137	0.144	0.149	0.150	0.153	0.153	0.158	0.164	0.183	✓	647
Mid-Cap Growth	0.424	1.45	0.148	0.158	0.167	0.174	0.179	0.183	0.188	0.194	0.198	0.214	✓	723
Small Value	0.465	1.50	0.125	0.141	0.146	0.150	0.153	0.157	0.160	0.163	0.169	0.187	✓	408
Small Blend	0.376	1.35	0.147	0.156	0.159	0.164	0.168	0.171	0.174	0.173	0.178	0.199	x	617
Small Growth	0.415	1.47	0.159	0.177	0.185	0.190	0.198	0.201	0.204	0.210	0.218	0.234	✓	596
Foreign Large Value	0.345	1.21	0.150	0.161	0.167	0.171	0.171	0.174	0.172	0.176	0.181	0.182	x	206
Foreign Large Blend	0.278	1.19	0.144	0.153	0.156	0.158	0.159	0.159	0.162	0.161	0.167	0.172	x	544
Foreign Large Growth	0.238	1.19	0.164	0.172	0.172	0.175	0.178	0.179	0.179	0.180	0.185	0.195	✓	245
Div. Emerging Mark.	0.310	1.24	0.188	0.203	0.212	0.215	0.216	0.220	0.222	0.224	0.222	0.233	x	351
World Stock	0.365	1.41	0.123	0.134	0.138	0.142	0.147	0.147	0.153	0.160	0.163	0.173	✓	555
Real Estate	0.313	1.29	0.140	0.168	0.168	0.170	0.171	0.169	0.175	0.175	0.176	0.181	x	133
Technology	0.422	1.45	0.187	0.216	0.228	0.235	0.232	0.234	0.240	0.241	0.252	0.272	x	165

Panel B: Persistence of ADD

Category	Persistence		Quantile Portfolios										incr.	# funds
	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10		
Large Value	0.330	1.43	0.035	0.038	0.040	0.041	0.041	0.042	0.043	0.044	0.045	0.050	✓	956
Large Blend	0.340	1.41	0.037	0.040	0.041	0.042	0.042	0.043	0.044	0.046	0.047	0.052	✓	1641
Large Growth	0.339	1.48	0.044	0.048	0.049	0.051	0.052	0.053	0.055	0.056	0.058	0.065	✓	1144
Mid-Cap Value	0.369	1.57	0.035	0.040	0.041	0.042	0.043	0.044	0.045	0.046	0.050	0.055	✓	456
Mid-Cap Blend	0.372	1.62	0.039	0.044	0.045	0.047	0.048	0.049	0.050	0.051	0.054	0.063	✓	647
Mid-Cap Growth	0.378	1.64	0.047	0.053	0.056	0.059	0.060	0.063	0.065	0.067	0.071	0.077	✓	723
Small Value	0.396	1.60	0.040	0.045	0.047	0.048	0.050	0.051	0.053	0.054	0.056	0.064	✓	408
Small Blend	0.311	1.48	0.048	0.051	0.052	0.054	0.055	0.056	0.056	0.057	0.059	0.071	✓	617
Small Growth	0.377	1.64	0.053	0.059	0.063	0.067	0.068	0.070	0.072	0.075	0.078	0.087	✓	596
Foreign Large Value	0.284	1.21	0.052	0.056	0.059	0.059	0.060	0.060	0.062	0.062	0.064	0.063	x	206
Foreign Large Blend	0.222	1.28	0.050	0.055	0.056	0.056	0.056	0.056	0.058	0.058	0.060	0.064	✓	544
Foreign Large Growth	0.201	1.22	0.058	0.060	0.061	0.062	0.064	0.066	0.066	0.065	0.067	0.071	x	245
Div. Emerging Mark.	0.284	1.27	0.070	0.076	0.080	0.081	0.082	0.083	0.084	0.085	0.085	0.089	✓	351
World Stock	0.288	1.59	0.041	0.045	0.047	0.049	0.051	0.051	0.053	0.056	0.057	0.065	✓	555
Real Estate	0.299	1.34	0.047	0.057	0.058	0.059	0.060	0.060	0.060	0.061	0.062	0.063	✓	133
Technology	0.373	1.49	0.071	0.081	0.086	0.088	0.088	0.089	0.093	0.093	0.101	0.106	✓	165

Table 3.2: Detailed drawdown persistence results for single fund categories. Only categories that contain more than 100 funds are included. All column variables are defined as in Table 3.1.

Panel A: Persistence of MDD

Time Period	$\mathcal{P}_{\text{cor}}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	incr.	# funds
	Persistence		Quantile Portfolios											
All Equity														
1990–1999	0.467	2.12	0.095	0.102	0.105	0.107	0.110	0.121	0.132	0.144	0.160	0.201	✓	2667
2000–2009	0.431	1.62	0.169	0.187	0.194	0.197	0.203	0.207	0.213	0.221	0.233	0.273	✓	5019
2010–2019	0.427	1.62	0.113	0.121	0.124	0.127	0.132	0.137	0.142	0.147	0.155	0.183	✓	4851
U.S. Equity														
1990–1999	0.538	2.00	0.091	0.096	0.100	0.102	0.107	0.110	0.122	0.136	0.154	0.182	✓	1811
2000–2009	0.444	1.48	0.170	0.184	0.188	0.191	0.196	0.202	0.208	0.214	0.226	0.252	✓	3388
2010–2019	0.455	1.46	0.110	0.116	0.120	0.121	0.125	0.130	0.134	0.139	0.144	0.161	✓	2924
Int. Equity														
1990–1999	0.440	2.00	0.110	0.112	0.116	0.118	0.123	0.124	0.130	0.154	0.177	0.220	✓	586
2000–2009	0.428	1.41	0.181	0.201	0.210	0.211	0.215	0.219	0.225	0.228	0.241	0.255	✓	1094
2010–2019	0.353	1.46	0.121	0.134	0.137	0.142	0.145	0.147	0.149	0.155	0.162	0.177	✓	1509
Sector Equity														
1990–1999	0.608	3.11	0.088	0.105	0.125	0.130	0.143	0.157	0.170	0.198	0.225	0.274	✓	266
2000–2009	0.467	2.12	0.157	0.186	0.210	0.215	0.212	0.225	0.246	0.271	0.296	0.333	x	544
2010–2019	0.418	2.32	0.115	0.129	0.138	0.144	0.145	0.151	0.152	0.170	0.196	0.267	✓	529

Panel B: Persistence of ADD

Time Period	$\mathcal{P}_{\text{cor}}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	incr.	# funds
	Persistence		Quantile Portfolios											
All Equity														
1990–1999	0.427	2.70	0.027	0.029	0.029	0.031	0.033	0.037	0.041	0.045	0.051	0.073	✓	2667
2000–2009	0.388	1.73	0.063	0.070	0.074	0.076	0.079	0.082	0.084	0.087	0.091	0.109	✓	5019
2010–2019	0.425	2.06	0.031	0.033	0.036	0.037	0.039	0.041	0.043	0.045	0.049	0.064	✓	4851
U.S. Equity														
1990–1999	0.487	2.24	0.025	0.026	0.027	0.028	0.030	0.033	0.037	0.041	0.046	0.056	✓	1811
2000–2009	0.399	1.60	0.062	0.068	0.070	0.073	0.075	0.078	0.082	0.084	0.089	0.099	✓	3388
2010–2019	0.467	1.75	0.028	0.030	0.031	0.032	0.034	0.036	0.038	0.040	0.042	0.049	✓	2924
Int. Equity														
1990–1999	0.422	2.28	0.036	0.036	0.036	0.037	0.039	0.041	0.045	0.058	0.066	0.082	✓	586
2000–2009	0.352	1.46	0.069	0.076	0.080	0.081	0.083	0.085	0.088	0.091	0.095	0.101	✓	1094
2010–2019	0.366	1.71	0.038	0.042	0.046	0.047	0.049	0.050	0.051	0.054	0.058	0.065	✓	1509
Sector Equity														
1990–1999	0.524	4.14	0.029	0.035	0.040	0.042	0.047	0.050	0.056	0.068	0.090	0.120	✓	266
2000–2009	0.397	2.23	0.060	0.073	0.082	0.082	0.082	0.087	0.101	0.116	0.124	0.134	✓	544
2010–2019	0.356	2.92	0.037	0.042	0.042	0.044	0.047	0.046	0.048	0.058	0.071	0.108	x	529

Table 3.3: Subperiod analysis of the persistence results. The sample period is split into three shorter periods (e.g., from January 1990 to December 1999). All variables are defined as in Table 3.1.

Measure	Length of t+1	Length of t−1 period								
		10 days	1 month	3 months	6 months	1 year	2 years	3 years	4 years	5 years
MDD	12 months	−0.141	−0.062	−0.034	−0.030	0.428	−0.001	−0.001	−0.036	−0.052
ADD	12 months	−0.182	−0.099	−0.065	−0.047	0.406	+0.015	+0.006	−0.019	−0.043
MDD	3 months	−0.056	+0.016	+0.029	+0.018	0.406	+0.006	+0.011	−0.036	−0.064
ADD	3 months	−0.070	−0.005	+0.011	+0.003	0.350	+0.015	+0.001	−0.020	−0.041
MDD	3 years	−0.172	−0.093	−0.051	−0.024	0.439	−0.007	−0.022	−0.039	−0.068
ADD	3 years	−0.206	−0.119	−0.070	−0.036	0.415	+0.004	−0.009	−0.040	−0.068

Table 3.4: Persistence for different lengths of the t−1 period between 10 days and five years. We compute the persistence measure $\mathcal{P}_{\mathrm{cor}}$ for all funds in the sample. For the one-year t−1 period, we provide the correlation $\mathcal{P}_{\mathrm{cor}}$ shaded in gray; for all other period lengths, we provide the difference to this value. In the first row, for example, the rank correlation of one-year MDD_{t+1} with one-year $\mathrm{MDD}_{t−1}$ is 0.428; when two years of t−1 data are used, the persistence with one-year MDD_{t+1} drops to 0.427, indicated in the table by −0.001. The rows differ in the length of the t+1 period as well as the measure MDD or ADD.

When an investor has decided that the annual MDD is of interest in the future, a second question regarding time becomes practically relevant: how much historical drawdown information shall be exploited? Intuitively, using one month of data may be insufficient, but using five years of data may be too long. To address this question, the length of the t−1 period is varied systematically, and the resulting persistence is documented. Table 3.4 illustrates how significantly changing the length of the t−1 period affects the results.[10] For example, when the past five years instead of the past one are used to predict the future 12-month MDD, the persistence drops by 0.052, from 0.428 to 0.376. If only one month of data is considered, the persistence drops by 0.062, from 0.428 to 0.366. The results from Table 3.4 show that when predicting 12-month MDD, one year of historical data yields the highest persistence. For the 12-month ADD in the second row, the persistence similarly decreases for quite long or short t−1 periods but increases slightly when two or three years of data are considered. When three months or three years are predicted in t+1, one might guess that the corresponding period length during t−1 might lead to the highest

---

[10]The results in Table 3.4 differ slightly from those in other tables. For example, Table 3.4 reports 0.428 for the persistence between annual $\mathrm{MDD}_{t−1}$ and annual MDD_{t+1}, while the corresponding result from Table 3.1 is 0.436. The difference is due to a slight adjustment in the simulation setup for Table 3.4 to create a level playing field for the comparison between different period lengths. While in Table 3.1 the first evaluated t+1 time window is year two because one year is necessary beforehand, the first evaluated time window in Table 3.4 is year six because the five-year t−1 window needs as much time beforehand. Hence, the year two evaluation is omitted in Table 3.4 to allow for a fair comparison between the one-year and five-year t−1 period lengths. As the adjustment is fairly small, the resulting differences are small as well.

persistence (i.e., three-month MDD_{t-1} is most correlated with three-month MDD_{t+1}). For the three-month predictions, this is indeed the case. However, the differences to the predictions based on one year of data are fairly small for both MDD and ADD. For the three-year predictions, using three years of data does not lead to improvements, but one year of data is optimal for the MDD and close to optimal for the ADD. In summary, the exact length of the $t-1$ period is not critically important for persistence, for example, the difference between using one year or two years of data is always negligible. Consequently, investors need not worry too much about how much history should be taken into account. However, unsurprisingly, using extremely short or extremely long data history significantly reduces persistence. Among all tested period lengths, the choice of one year of historical data is always either optimal or close to optimal. Therefore, regardless of the $t+1$ period, we adhere to a period length of one year during $t-1$ for simplicity unless indicated otherwise.

In conclusion, mutual fund drawdown is persistent. To support this claim, we have systematically investigated two drawdown measures, MDD and ADD, for numerous sets of mutual funds from various fund categories using two different persistence measures, $\mathcal{P}_{cor}$ and $\mathcal{P}_{10/1}$. Further robustness checks regarding the period of investigation and length of the $t-1$ and $t+1$ periods confirm that drawdown persistence is robust.

3.5 Predicting Drawdown with the Volatility Shortcut

Thus far, extensive evidence has been collected that past MDD and ADD are persistent, i.e., they are informative in predicting future MDD and ADD. Therefore, consulting past MDD or ADD to estimate future drawdown is reasonable. However, better options for predicting future drawdown may be available. While any arbitrary measure or quantity might theoretically serve as a predictor, we focus on a ubiquitous and straightforward risk measure: the historical standard deviation of returns.[11] This choice is not too far-fetched for two reasons. First, as discussed in the literature review, volatility itself is highly persistent. Second, drawdown measures are sensitive to volatility because large fluctuations in prices usually incur high drawdowns. Therefore, historical standard deviation may contain information about future drawdown.

---

[11] Along with the standard deviation, we have also analyzed other measures to predict future drawdown, for example, realized 10-day and one-year value-at-risk at the 95% confidence level and momentum-based estimators. Of these, the standard deviation performed best. However, we do not claim that the standard deviation is the best possible predictor because we have not attempted an all-encompassing search to find the optimal predictor.

How accurately the drawdown measures predict themselves has been answered in Section 3.4; to compare these results with the standard deviation, this analysis requires only slight adaptation. Instead of asking MDD_{t-1} to predict MDD_{t+1}, we ask the standard deviation during $t-1$ (henceforth denoted by StdDev_{t-1}) to predict MDD_{t+1}. All statistics for evaluating persistence, such as the measures $\mathcal{P}_{\text{cor}}$ and $\mathcal{P}_{10/1}$, can be computed analogously for the standard deviation. The analogous statistic, which compares two different risk measures, can be interpreted just as well as the persistence statistic, where the same risk measure is applied both in $t-1$ and $t+1$. For the entire sample, the results show that $\mathcal{P}_{\text{cor}}$ of the standard deviation predicting MDD_{t+1} exceeds $\mathcal{P}_{\text{cor}}$ of the drawdown measure itself by 0.493 to 0.436. Similarly, $\mathcal{P}_{10/1}$ yields higher persistence for the portfolios sorted by standard deviation (1.90) compared to those sorted by MDD_{t-1} (1.74). These results hold for all category groups and when the ADD is substituted for the MDD, see Table 3.5 for all details and Tables 3.15 and 3.16 in the appendix for different period lengths. The differences between standard deviation and drawdown are rather small for international funds but sizable for U.S. and sector funds.

| Category Group | Persistence | | Quantile Portfolios | | | | | | | | | | |
	$\mathcal{P}_{\text{cor}}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	incr.
MDD:													
All Equity	0.493	1.90	0.123	0.137	0.141	0.145	0.150	0.155	0.162	0.172	0.188	0.234	✓
	0.436	1.74	0.127	0.139	0.143	0.146	0.150	0.157	0.164	0.173	0.185	0.221	✓
U.S. Equity	0.563	1.77	0.119	0.131	0.135	0.139	0.143	0.148	0.156	0.167	0.182	0.211	✓
	0.465	1.61	0.125	0.134	0.138	0.140	0.145	0.149	0.157	0.165	0.177	0.201	✓
Int. Equity	0.433	1.66	0.134	0.148	0.154	0.160	0.164	0.167	0.172	0.182	0.195	0.222	✓
	0.393	1.58	0.139	0.151	0.157	0.159	0.163	0.165	0.170	0.181	0.195	0.219	✓
Sector Equity	0.609	2.72	0.110	0.125	0.145	0.158	0.169	0.182	0.198	0.230	0.262	0.299	✓
	0.468	2.40	0.122	0.142	0.160	0.165	0.169	0.179	0.192	0.215	0.242	0.293	✓
ADD:													
All Equity	0.451	2.27	0.040	0.045	0.046	0.048	0.050	0.052	0.055	0.059	0.066	0.091	✓
	0.409	2.02	0.041	0.045	0.047	0.049	0.051	0.055	0.057	0.060	0.065	0.083	✓
U.S. Equity	0.528	2.08	0.037	0.041	0.042	0.044	0.047	0.048	0.052	0.057	0.063	0.077	✓
	0.441	1.77	0.039	0.043	0.044	0.046	0.048	0.050	0.053	0.056	0.060	0.069	✓
Int. Equity	0.419	1.96	0.045	0.051	0.054	0.057	0.058	0.060	0.063	0.068	0.075	0.088	✓
	0.369	1.75	0.048	0.052	0.055	0.056	0.058	0.060	0.063	0.068	0.074	0.084	✓
Sector Equity	0.544	3.63	0.035	0.041	0.050	0.055	0.059	0.064	0.071	0.089	0.106	0.127	✓
	0.400	2.88	0.042	0.051	0.056	0.057	0.060	0.062	0.070	0.083	0.096	0.121	✓

Table 3.5: Analysis of how the standard deviation predicts drawdown. The results based on the standard deviation are printed in black; the previous persistence results from Table 3.1 are displayed in gray. All variables are defined as in Table 3.1.

For an in-depth look, the predictive power of MDD_{t-1}, ADD_{t-1}, and $StdDev_{t-1}$ is also compared for single fund categories. To this end, we repeat the computations leading to Table 3.2 analogously for the standard deviation. The resulting Table 3.6 can then be compared to the drawdown-based Table 3.2. For an easier comparison of the two tables, see Figure 3.2. The standard deviation predicts MDD_{t+1} and ADD_{t+1} better than the corresponding drawdown measures in every single category and based on every metric. While the measure $\mathcal{P}_{cor}$ for the standard deviation exceeds the measure for MDD and ADD by a fairly uniform margin across all categories, the differences in $\mathcal{P}_{10/1}$ vary. For example, the difference between the standard deviation and the drawdown measures is more pronounced in the U.S. categories than in the international categories. Besides $\mathcal{P}_{cor}$ and $\mathcal{P}_{10/1}$, a comparison of single decile portfolios yields that the portfolio $\mathcal{Q}_1$ sorted on lowest $StdDev_{t-1}$ always has lower MDD_{t+1} than the portfolio sorted on lowest MDD_{t-1}; similarly, the portfolio $\mathcal{Q}_{10}$ sorted on highest $StdDev_{t-1}$ always yields higher MDD_{t+1} than the portfolio sorted on highest MDD_{t-1}.

In summary, we find that the standard deviation is more informative about future drawdown than the drawdown measures themselves. Thus, past standard deviation is the superior indicator when the future drawdown shall be managed. Without computing a single drawdown metric, practitioners may take the shortcut of using past volatility to manage future drawdown. The potential implications of this finding are significant. Even investors who actively seek to minimize future drawdown should choose funds with low past volatility instead of funds with low historical drawdown. Consequently, reporting past MDD or ADD in fund prospectuses would provide few benefits. Conversely, investors who choose low volatility funds might already implicitly enjoy relatively low future drawdown. Hence, low volatility funds may add low expected drawdowns to their selling points. As the standard deviation is well-established and often computed and reported already, most implications would remain valid even if – hypothetically – the standard deviation predicted drawdown only equally reliably but not significantly better. Although some link between volatility and drawdown should be expected, it may come as a surprise that volatility's predictive power exceeds that of the drawdown measure itself. Possible explanations for the results are discussed in depth in the following section.

Panel A: Persistence of MDD

Category	Persistence		Quantile Portfolios										incr.	# funds
	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10		
Large Value	0.489	1.45	0.107	0.119	0.125	0.130	0.132	0.135	0.136	0.138	0.143	0.155	✓	956
Large Blend	0.417	1.41	0.113	0.126	0.131	0.131	0.132	0.133	0.136	0.138	0.144	0.159	✓	1641
Large Growth	0.512	1.53	0.127	0.138	0.143	0.148	0.152	0.157	0.161	0.166	0.175	0.194	✓	1144
Mid-Cap Value	0.509	1.48	0.114	0.126	0.132	0.137	0.138	0.142	0.146	0.148	0.151	0.169	✓	456
Mid-Cap Blend	0.518	1.61	0.121	0.134	0.142	0.145	0.150	0.152	0.156	0.159	0.164	0.195	✓	647
Mid-Cap Growth	0.521	1.58	0.142	0.154	0.164	0.171	0.175	0.184	0.190	0.194	0.203	0.225	✓	723
Small Value	0.529	1.58	0.120	0.138	0.146	0.150	0.154	0.158	0.161	0.164	0.169	0.190	✓	408
Small Blend	0.461	1.41	0.144	0.151	0.157	0.163	0.168	0.170	0.173	0.177	0.184	0.203	✓	617
Small Growth	0.514	1.61	0.152	0.169	0.183	0.190	0.194	0.203	0.208	0.213	0.220	0.245	✓	596
Foreign Large Value	0.416	1.23	0.150	0.161	0.165	0.167	0.169	0.170	0.175	0.180	0.183	0.185	✓	206
foreign Large Blend	0.325	1.23	0.141	0.151	0.155	0.158	0.160	0.160	0.162	0.165	0.166	0.174	✓	544
Foreign Large Growth	0.308	1.20	0.163	0.168	0.172	0.173	0.175	0.181	0.182	0.185	0.185	0.196	✓	245
Div. Emerging Mark.	0.405	1.26	0.184	0.202	0.209	0.216	0.217	0.221	0.224	0.225	0.227	0.232	✓	351
World Stock	0.414	1.49	0.120	0.132	0.137	0.143	0.145	0.149	0.151	0.159	0.166	0.179	✓	555
Real Estate	0.375	1.37	0.134	0.169	0.167	0.172	0.171	0.173	0.171	0.174	0.177	0.183	x	133
Technology	0.524	1.54	0.183	0.213	0.218	0.226	0.229	0.234	0.242	0.251	0.258	0.282	✓	165

Panel B: Persistence of ADD

Category	Persistence		Quantile Portfolios										incr.	# funds
	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10		
Large Value	0.433	1.58	0.033	0.037	0.039	0.041	0.041	0.042	0.043	0.044	0.046	0.052	✓	956
Large Blend	0.355	1.53	0.036	0.040	0.042	0.042	0.042	0.042	0.043	0.045	0.048	0.055	✓	1641
Large Growth	0.458	1.71	0.041	0.045	0.047	0.049	0.051	0.053	0.055	0.057	0.061	0.070	✓	1144
Mid-Cap Value	0.472	1.68	0.034	0.039	0.040	0.042	0.043	0.045	0.046	0.047	0.048	0.057	✓	456
Mid-Cap Blend	0.460	1.82	0.038	0.041	0.045	0.045	0.047	0.049	0.050	0.052	0.054	0.069	✓	647
Mid-Cap Growth	0.481	1.87	0.045	0.049	0.053	0.056	0.060	0.064	0.066	0.068	0.073	0.084	✓	723
Small Value	0.482	1.71	0.038	0.044	0.047	0.049	0.050	0.051	0.054	0.054	0.056	0.065	✓	408
Small Blend	0.400	1.61	0.046	0.048	0.050	0.053	0.054	0.055	0.057	0.058	0.062	0.074	✓	617
Small Growth	0.477	1.92	0.049	0.056	0.061	0.064	0.066	0.071	0.074	0.076	0.080	0.094	✓	596
Foreign Large Value	0.375	1.33	0.051	0.055	0.057	0.058	0.059	0.060	0.062	0.063	0.064	0.068	✓	206
Foreign Large Blend	0.299	1.35	0.048	0.053	0.055	0.056	0.057	0.057	0.058	0.059	0.060	0.065	✓	544
Foreign Large Growth	0.280	1.30	0.056	0.059	0.061	0.062	0.062	0.065	0.067	0.067	0.068	0.073	✓	245
Div. Emerging Mark.	0.389	1.30	0.069	0.074	0.078	0.081	0.081	0.084	0.085	0.086	0.087	0.090	✓	351
World Stock	0.350	1.74	0.039	0.044	0.046	0.049	0.050	0.052	0.052	0.056	0.060	0.068	✓	555
Real Estate	0.369	1.42	0.045	0.058	0.057	0.060	0.060	0.061	0.059	0.061	0.063	0.064	x	133
Technology	0.507	1.79	0.066	0.077	0.080	0.084	0.086	0.089	0.093	0.099	0.102	0.118	✓	165

Table 3.6: Detailed persistence results for single fund categories as in Table 3.2 but with the standard deviation applied during t−1. During t+1, MDD$_{t+1}$ is applied in Panel A, and ADD$_{t+1}$ is applied in Panel B.
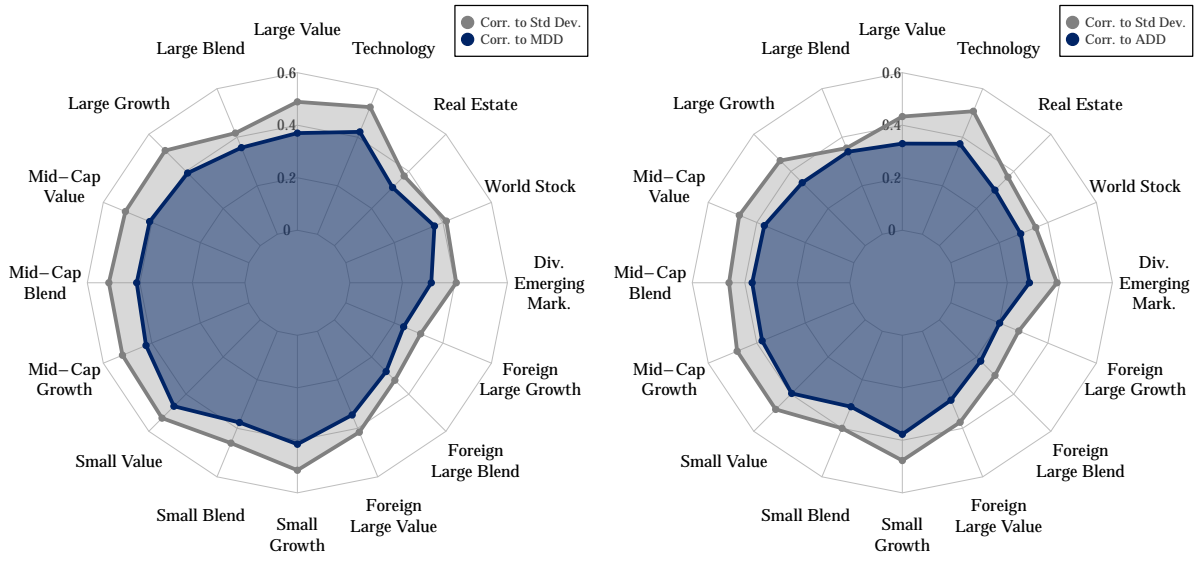
(a) Comparison of $\mathcal{P}_{\mathrm{cor}}$ when predicting MDD_{t+1}.

(b) Comparison of $\mathcal{P}_{\mathrm{cor}}$ when predicting ADD_{t+1}.
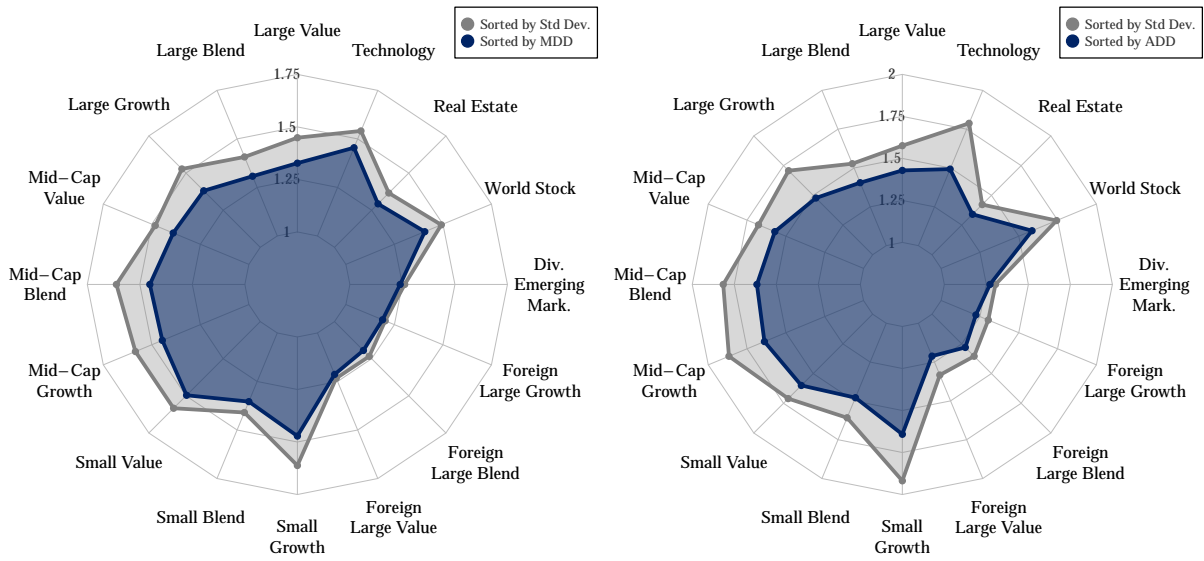
(c) Comparison of $\mathcal{P}_{10/1}$ when predicting MDD_{t+1}.

(d) Comparison of $\mathcal{P}_{10/1}$ when predicting ADD_{t+1}.

Figure 3.2: Summary of the prediction of MDD_{t+1} or ADD_{t+1} with the standard deviation compared to the drawdown measure itself across fund categories. The drawdown measure is depicted in light gray while the standard deviation is depicted in dark blue. Plots (a) and (b) use persistence measure $\mathcal{P}_{\mathrm{cor}}$ for the comparison, and plots (c) and (d) use $\mathcal{P}_{10/1}$. In plots (a) and (c), MDD_{t+1} is the target measure; in plots (b) and (d), it is ADD_{t+1}. All plots result from a comparison of results from Tables 3.2 and 3.6.

3.6 The Role of Volatility and Return Persistence

In the previous section, we have collected plenty of evidence for a surprising finding: past standard deviation is a more accurate indicator of future drawdown than the drawdown measure itself. In this section, we provide an explanation for this puzzling result, which evolves around volatility persistence and the (non-)existence of persistence in returns. In the following, we first discuss data-based evidence and then provide more details with a tailor-made simulation study.

The data-based approach elaborates on the previous correlation analysis, in which rank correlations of MDD_{t-1} and MDD_{t+1} as well as StdDev_{t-1} and MDD_{t+1} have been reported. Now, correlations for all possible combinations of $t-1$ and $t+1$ quantities are computed for the drawdown measures, volatility, and returns. For simplicity, we first focus on annual MDD. When correlations for quantities within the same period are computed as in the left portion of Table 3.7, the MDD is highly positively correlated with the standard deviation (0.677) and fairly negatively correlated with the return (-0.346). Both results are intuitive because drawdowns tend to increase when prices are volatile and they tend to decrease when prices rise. In the right portion of Table 3.7, we report correlations between consecutive periods, i.e., when one measure ranks during period $t-1$ and the other measure ranks during period $t+1$. As before, we observe the puzzling fact that the relationship between past and future MDD is weaker (0.436) than the relation between past standard deviation and future MDD (0.493). Also apparent is the stark difference between return and standard deviation regarding their respective persistence properties. While the standard deviation is strongly persistent (0.720), the funds in the sample exhibit hardly any return persistence (0.103). Both findings are plausible as they are consistent with the literature, cf. Section 3.2.

Do the persistence results for the return and the standard deviation help to explain our puzzle? They do, if you consider the following explanation: at the end of period $t-1$, we would like to make an educated guess about the future MDD. Although past returns have been significantly correlated with past MDD, considering past returns to predict future MDD does not seem like a good idea because the return information itself is not particularly persistent (0.103). Consequently, past return provides no information about future MDD, and correlations are close to zero (0.030). Things are different for the standard deviation. The information contained in the standard deviation is persistent across time (0.720) and highly correlated with MDD in the same period (0.677). Consequently, past standard deviation is a useful tool for predicting future MDD. But what about past MDD? It is affected by both past standard deviation and return, of which

Panel A: Correlations of relative ranks with MDD

t−1	Return	Std Dev.	MDD	Return	Std Dev.	MDD
		t−1			t+1	
Return	1	0.013	−0.346	0.103	0.088	0.030
Std Dev.		1	0.677	−0.003	0.720	0.493
MDD			1	−0.032	0.482	0.436

Panel B: Correlations of relative ranks with ADD

t−1	Return	Std Dev.	ADD	Return	Std Dev.	ADD
		t−1			t+1	
Return	1	0.013	−0.508	0.103	0.088	−0.030
Std Dev.		1	0.592	−0.003	0.720	0.451
ADD			1	−0.070	0.416	0.409

Table 3.7: Correlations between different measures within the period or with the consecutive period. For example, fund ranks based on return and MDD of the same period (t−1) are correlated −0.346; ranks based on return during t−1 and MDD during t+1 are correlated 0.030. All correlations are based on the full data set and correspond to the persistence measure $\mathcal{P}_{\mathrm{cor}}$. For a full correlation matrix, copy the top left submatrix in the bottom right submatrix, since correlations for the same subperiod are the same, irrespective of it being called t−1 or t+1. Since it carries no additional information, it has been omitted.

one is useful while the other is not. In a sense, the useful information that the MDD contains from the standard deviation is obstructed by noise from the return information that is not useful for predicting future MDD. Therefore, it is reasonable that the standard deviation alone is a better predictor of future MDD. In addition to the annual correlation results, all correlations for periods of three months and three years support this argument, cf. Tables 3.17 and 3.18.

If this explanation were true, we would expect that when a fund's return persistence is high, its drawdown persistence is high, too. For each fund and point in time, we examine whether such a relationship exists in the mutual fund data. We quantify the return persistence of fund i between times t−1 and t+1 by one minus the absolute difference between the relative rank of the fund's return during t−1 and its relative rank during t+1, such that persistence is low when relative ranks are far apart.[12] Drawdown persistence is quantified analogously. For example, when a fund has the highest MDD during t−1 among 1,000 funds and then ranks second according to

---

[12]To quantify persistence here, we cannot use $\mathcal{P}_{\mathrm{cor}}$ or $\mathcal{P}_{10/1}$ because we need a persistence measure for a single fund and a single point in time.

MDD during t+1, its persistence is $1 - \left| \frac{1}{1000} - \frac{2}{1000} \right| = 0.999$, quite close to perfect persistence. If the formerly highest MDD fund exhibited the second-lowest MDD during t+1, we would quantify persistence as $1 - \left| \frac{1}{1000} - \frac{999}{1000} \right| = 0.002$, i.e., quite low. To determine whether there is a relationship between return persistence and MDD persistence, the correlation coefficient is computed for all pairs of return and MDD persistence for each fund and each period. We obtain a correlation of 0.21, indicating that return persistence and MDD persistence are indeed positively correlated.

Similarly, all arguments hold for the average drawdown, see Table 3.7. For the relation between return persistence and ADD persistence, an even stronger positive correlation of 0.37 is obtained. All results for both ADD and MDD hold similarly for all category groups.

Simulating Persistence

The analysis of our data set of mutual funds reveals that drawdown persistence is intricately linked to volatility and return persistence. Volatility persistence emerges as the main driver of drawdown persistence, but when return persistence is present, it adds significantly to drawdown persistence. Are these results specific to mutual funds or can they be generalized? To address whether the previous results are caused by idiosyncratic particularities in the mutual fund data, we develop a suitable simulation model. Its purpose is to model what happens to drawdown persistence when return or volatility persistence increase. Therefore, return persistence and volatility persistence serve as inputs to the model, and we observe the drawdown persistence of the output.

In general, simulating drawdown requires discretized sample paths, which is why we model daily log return time series. At large, the model has a stochastic mean stochastic variance structure to allow return and volatility to be simulated separately. Both stochastic processes for the mean and volatility component are fractionally integrated (FI) processes introduced by Granger (1980) because they are ideal to model persistence. Compared to autoregressive moving-average (ARMA) processes, which are designed to model short-range dependence, FI processes excel at modeling long-range dependence, which coincides with our notion of persistence. The combination of both classes of models, autoregressive fractionally integrated moving average (ARFIMA) processes, have been studied extensively (see Bhardwaj and Swanson (2006), and the references therein) and have been applied in the finance context (e.g., Caporale and Gil-Alana (2006) and Aye et al. (2014)). Since we focus on simulating persistence and, therefore, long-range dependence, FI

processes are used to model each component of the stochastic mean stochastic variance model. This approach allows us to efficiently model both return and volatility persistence.

We model daily log returns in a stochastic mean stochastic variance model with fractionally integrated subprocesses. The returns r_t are simulated via

$$r_t = m_t + \sqrt{v_t}\, \varepsilon_t \qquad \text{with} \qquad m_t = \text{FI}(d_1, \mu_1, \sigma_1)$$

$$v_t = \exp(\tilde{v}_t)$$

$$\tilde{v}_t = \text{FI}(d_2, \mu_2, \sigma_2),$$

where m_t is the stochastic mean component, v_t is the stochastic volatility component, and $\varepsilon_t \sim \mathcal{N}(0,1)$. The fractionally integrated processes $\{X_t\}_{t=0,\dots,N} = \text{FI}(d, \mu, \sigma)$ are defined via

$$(1-B)^d X_t = \eta_t \qquad \text{with } \eta_t \sim \mathcal{N}(\mu, \sigma),$$

where B is the backshift operator, i.e., $BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$ etc., cf. Granger (1980). What makes the FI process 'fractional' is that the exponent d is not necessarily an integer but a rational number, usually strictly between 0 and 0.5 to ensure positive dependence and stationarity (Hosking, 1981). To handle such expressions, the binomial series can be applied to yield

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k$$

using generalized binomial coefficients defined as

$$\binom{d}{k} = \frac{d(d-1) \cdots (d-k+1)}{k!}.$$

The infinite series would terminate when k exceeds d if the exponent d were an integer. Compared to an ARMA model, the FI model's autocorrelation function exhibits slow hyperbolic decay instead of exponential decay (Baillie, 1996). Due to the slower decay, the FI processes are long-range dependent, which allows us to model persistence. Intuitively, the fractionally integrating parameters d_1 and d_2 steer the return and volatility persistence, respectively, whereas higher values of d_1 or d_2 correspond to stronger persistence.

Four scenarios are considered to analyze the impact of return and volatility persistence: only the return being persistent, only the standard deviation being persistent, both being persistent, or neither. In each scenario, 100,000 paths of daily returns are simulated for two years and the correlation-based persistence measure $\mathcal{P}_{\text{cor}}$ is computed between years one and two. As mentioned previously, return and volatility persistence can be steered using the parameters d_1 and d_2, respectively. If neither should be persistent, we set $d_1 = d_2 = 0$. If only volatility should

be persistent, we choose $d_1 = 0$ and $d_2 = 0.42$ such that the simulated volatility persistence coincides with the volatility persistence of the mutual fund data in Table 3.7. To simulate a scenario without volatility persistence but with strong return persistence, we choose $d_2 = 0$ and $d_1 = 0.25$ resulting in a return persistence of 0.32, which is strong compared to 0.10 in Table 3.7. When both return and volatility should persist, both non-zero choices of d_1 and d_2 are used accordingly.

To obtain realistic choices for the remaining parameters, we compute daily log returns of the mutual fund data. The average daily return is 0.000261 (annualized 6.55%), and the standard deviation is 0.01278 (annualized 20.20%). As μ_1 directly determines the daily return of the process, we set it to 0.000261. To obtain the desired standard deviation, we set σ_1, σ_2, and μ_2 as follows. Observe that $\tilde{v}_t$ is the true variance process that is transformed with the exponential function to yield v_t, thus ensuring that the variance process is strictly positive. Reasonably sized volatility $\tilde{v}_t$ after the transformation is obtained by setting μ_2 to -10. Next, total volatility is split such that the volatility process contributes 75% and the mean process contributes 25%.[13] Moreover, the choice of d_1 or d_2 for each scenario slightly affects the standard deviation because higher persistence induces higher volatility. We correct for this effect by choosing σ_1 and σ_2 such that each process contributes the required amount to the total standard deviation.

The results for the four scenarios strongly support the previous findings on the drivers of drawdown persistence and are illustrated in Figure 3.3. For each scenario, the figure contrasts how MDD_{t-1} predicts MDD_{t+1} (henceforth called MDD persistence) against how StdDev_{t-1} predicts MDD_{t+1} (henceforth called SD-MDD persistence), analogously for ADD instead of MDD. First, return and volatility persistence are strongly underscored as drivers of drawdown persistence because, in the scenario without return and volatility persistence, there is neither MDD persistence nor SD-MDD persistence. This result also suggests that higher-order effects do not play a significant role. Second, return persistence sharply increases MDD persistence but ignores SD-MDD persistence. This result demonstrates a strong positive link between return persistence and MDD persistence. While the mutual fund data set also suggests this relationship, the simulation result here is more straightforward to interpret. In the third scenario, with volatility persistence only, the MDD is strongly persistent, but SD-MDD persistence is stronger by a significant margin. This result supports the evidence of volatility persistence being the driver of both MDD and SD-MDD

---

[13]Splitting total volatility is straightforward in our model, as the mean process m_t and the volatility process v_t are stochastically independent.

persistence. Naturally, SD-MDD persistence is affected more strongly than MDD persistence when the standard deviation itself is persistent. In the fourth scenario, both return and standard deviation are persistent. As MDD persistence exceeds SD-MDD persistence, the additional return persistence (which promotes MDD persistence) compared to the third scenario appears to turn the tide in favor of MDD persistence. The higher return persistence seems to dilute the volatility information and, thus, cause lower SD-MDD persistence compared to the third scenario. The right-hand side of Figure 3.3 contains the corresponding results for the ADD, which are qualitatively similar. Relatively, the effect of the return persistence is stronger here compared to the effect of the volatility persistence. Congruently, the return effect dominates in the fourth scenario, driving ADD persistence further ahead of SD-ADD persistence.
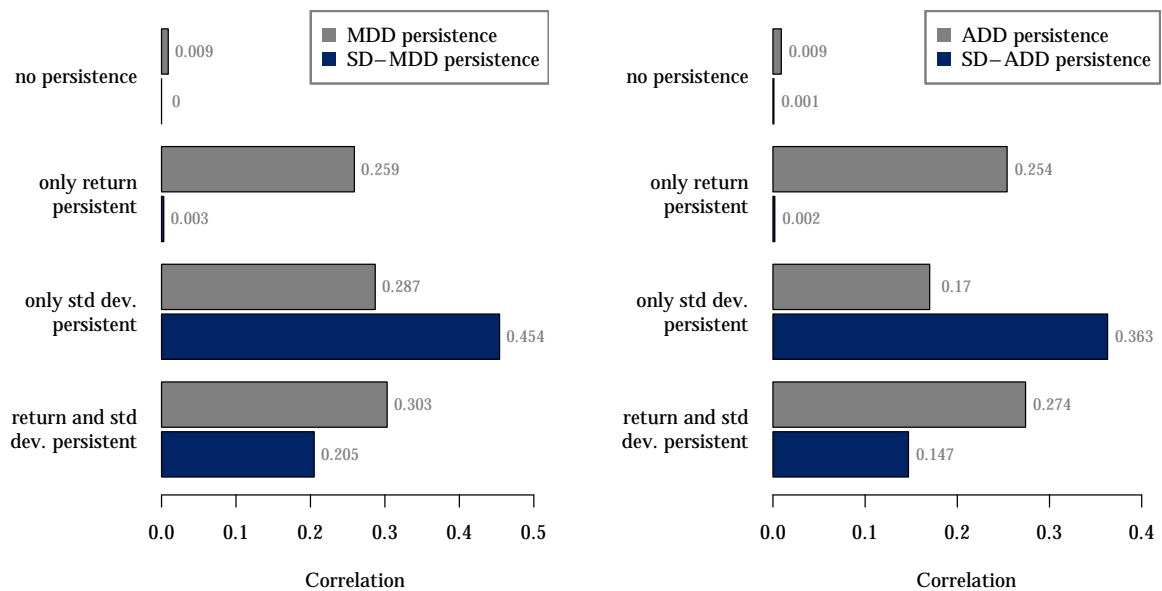
Figure 3.3: Summary of the effect which persistence in returns and standard deviation exert on drawdown persistence. The four scenarios – only return persistence, only standard deviation persistence, neither, or both – differ only in the prescribed model parameters for the corresponding persistence. From the simulated time series, the persistence between MDD_{t-1} and MDD_{t+1} (displayed in gray) and the persistence between $StdDev_{t-1}$ and MDD_{t+1} (displayed in blue) are computed using the correlation of relative ranks, i.e., the persistence measure $\mathcal{P}_{cor}$. Substitute MDD with ADD to obtain the right set of bar plots.

How do the simulated results fit to the results from the mutual fund data? To relate both analyses, recall that in the mutual fund data the standard deviation is much more persistent (rank correlation between $t-1$ and $t+1$ of 0.720) than the return (0.103). Hence, the fund data relates best to the third simulated scenario with a slight tilt toward scenario four. Therefore, we would expect from the simulations that MDD persistence trails SD-MDD persistence slightly, which is exactly what Table 3.7 reports.

In summary, the simulations support many interesting results. Return and volatility persistence appear to drive MDD and SD-MDD persistence (all analogous for the ADD), whereas return persistence only affects MDD persistence, and volatility persistence affects both but favors SD-MDD persistence. In settings in which the volatility persistence is dominant (as in scenario three or the mutual fund data), the standard deviation is the better MDD predictor. When strong return persistence is present in scenario four, the MDD predicts more accurately. Hence, the choice of the most adequate drawdown predictor depends on assumptions regarding return persistence. When investors feel confident in assuming high return persistence, relying on drawdown persistence may be warranted. If return persistence is largely absent, going for the standard deviation may be advantageous.

3.7 Does Managing Drawdowns Sacrifice Returns?

The empirical results virtually serve on a silver plate a straightforward strategy with which to manage future drawdown risk: if low future drawdown is desired, choose funds with low drawdown or low volatility in the past; if high future drawdown is desired, choose funds with high historical drawdown or volatility. What are the implications of pursuing such a strategy? In practice, the first question would regard returns: does managing drawdown risk lead to lower returns? Addressing this question is relevant because, depending on this relationship, our previous results have a different flavor. If low[14] future drawdown goes hand in hand with low returns, investors have to weight any decrease in future drawdown risk with a potential reduction in returns. If managing drawdown were to substantially sacrifice returns, managing drawdown might be impractical altogether. Conversely, if low drawdown coincided with higher returns, managing drawdown would be even more appealing in practice.

Before turning to the empirical analysis, let us consider what relationship may be expected. On the one hand, the usual risk–return trade-off, where higher risk is compensated with higher returns, suggests that funds with low drawdown should exhibit low returns. On the other hand, low-risk anomalies have been documented, for example, in Haugen and Heins (1975), Ang et al. (2006), Baker et al. (2011), and Frazzini and Pedersen (2014), where the effect is reversed. Regarding the volatility shortcut, the low volatility anomaly specifically suggests the contrary relationship between past standard deviation and future returns. While the empirical evidence is

---

[14]Of course, building portfolios with low drawdown is not always desirable. Instead, the goal is usually to steer the drawdown level depending on the individual risk appetite. We choose the case of low drawdown for illustrative purposes because we deem it most intuitive.

usually based on stocks, a similar anomaly for mutual funds may be hypothesized. With regard to exploiting drawdown persistence, support for a reverse relationship arises from the fact that drawdown is not ignorant of returns; in fact, high returns often coincide with low drawdown (see Table 3.7) such that maneuvering an investment toward low drawdown should also coincide with higher returns. Considering these conflicting arguments, relationships in both directions are conceivable.

Therefore, we leverage the mutual fund data set to address the relationship empirically. Instead of the persistence question, i.e., how MDD_{t-1} affects MDD_{t+1}, we observe how MDD_{t-1} affects average daily log returns during $t+1$ (analogously for ADD_{t-1} or $StdDev_{t-1}$). The results in Table 3.8 suggest that there is no systematic relationship between managing drawdown and future return. All correlations $\mathcal{P}_{cor}$ between MDD_{t-1} and subsequent returns are close to zero. Similarly, the average returns of the quantile portfolios neither monotonically increase nor decrease (except for international funds sorted on ADD or standard deviation). If at all, there is evidence that managing drawdown leads to slightly higher returns during $t+1$ because all correlations $\mathcal{P}_{cor}$ are slightly negative, and returns more often decrease than increase for higher deciles. For sector equity, the smallest category group, historical drawdowns in the highest deciles strongly signal lower future returns. Overall, managing drawdowns does not sacrifice returns. These results hold for managing drawdown with MDD_{t-1}, ADD_{t-1}, and $StdDev_{t-1}$ (see Table 3.8) as well as for three-month and three-year periods (see Table 3.19 in the appendix).

Even in the absence of positive or negative returns as a consequence of managing drawdown, high or low drawdown funds differ in risk; therefore, they potentially earn different α with respect to a factor model. For example, despite having similar returns, high drawdown funds may earn higher α than low drawdown funds. Similarly, high and low drawdown funds may differ in how much of their return can be attributed to the different factors. For instance, in comparison to low drawdown funds, high drawdown funds may contain more stocks of small firms, such that more of their excess return is explained by the size factor. To investigate both points, Fama-French 5-factor alphas are computed by regressing the mutual fund returns on the factor portfolios of the Fama-French 5-factor (FF5) model.[15] The necessary daily return time series of the five factors

---

[15]The FF5 factor model consists of five factor portfolios: the value-weighted market portfolio in excess of the risk-free rate, the SMB long–short portfolio of small minus big size stocks, the HML long–short portfolio of high minus low book-to-market stocks, the RMW long–short portfolio of robust minus weak operating profitability stocks, and the CMA long–short portfolio of conservative minus aggressive investment stocks (Fama and French, 2015).

| | Persistence | Quantile Portfolios | | | | | | | | | | | |
Category Group	$\mathcal{P}_{cor}$	1	2	3	4	5	6	7	8	9	10	incr.	decr.
MDD:													
All Equity	−0.032	0.078	0.076	0.078	0.082	0.081	0.079	0.077	0.077	0.074	0.052	x	x
U.S. Equity	−0.002	0.081	0.082	0.082	0.084	0.086	0.085	0.085	0.086	0.084	0.073	x	x
Int. Equity	−0.055	0.074	0.069	0.065	0.063	0.059	0.057	0.052	0.049	0.043	0.044	x	x
Sector Equity	−0.082	0.097	0.094	0.078	0.084	0.092	0.096	0.090	0.067	0.033	−0.012	x	x
ADD:													
All Equity	−0.070	0.083	0.082	0.081	0.081	0.080	0.078	0.075	0.073	0.071	0.049	x	x
U.S. Equity	−0.043	0.087	0.086	0.085	0.086	0.083	0.083	0.083	0.084	0.081	0.071	x	x
Int. Equity	−0.071	0.077	0.070	0.066	0.063	0.059	0.055	0.053	0.047	0.047	0.041	x	✓
Sector Equity	−0.108	0.097	0.095	0.091	0.099	0.092	0.088	0.082	0.055	0.030	−0.008	x	x
Std Dev.:													
All Equity	−0.003	0.078	0.076	0.077	0.077	0.076	0.075	0.077	0.079	0.079	0.059	x	x
U.S. Equity	−0.017	0.083	0.086	0.085	0.084	0.082	0.084	0.084	0.083	0.081	0.077	x	x
Int. Equity	−0.061	0.079	0.070	0.063	0.060	0.056	0.056	0.054	0.050	0.048	0.040	x	✓
Sector Equity	−0.080	0.095	0.093	0.077	0.087	0.093	0.093	0.085	0.053	0.028	0.017	x	x

Table 3.8: Effect of managing drawdown on fund returns. Sorting funds on MDD_{t-1}, ADD_{t-1}, or $StdDev_{t-1}$, we report rank correlations with future return ($\mathcal{P}_{cor}$) and average annual returns for each quantile portfolio. As trends across quantile portfolios are mostly absent, we omit computing $\mathcal{P}_{10/1}$. Instead, we add a column that indicates whether returns are monotonically decreasing.

are obtained from Kenneth French's website.[16] In detail, we first sort all funds by MDD_{t-1} into decile portfolios before we regress each fund's excess returns during t+1 on the returns of the FF5 factors during t+1.[17] We collect the resulting factor loadings β_i and intersect α for each fund as well as the factor risk premium for each period and factor. The product of factor loading β_i and the average factor return can be interpreted as the contribution of the factor to the average return. The resulting factor contributions and alphas are averaged across funds and time with one-year rolling windows at a monthly frequency.[18] To allow for a better interpretation, the daily results are annualized to arrive at Table 3.9 and its visualization in Figure 3.4.

In general, all alphas are fairly close to zero and mostly negative. Across decile portfolios, there are no significant trends to be observed. Hence, future α is largely unaffected by managing drawdown. Regarding factor contributions, the market factor β has the highest contribution

---

[16] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (last access: 31.03.2020)

[17] We restrict our analysis to the U.S. data set because the FF5 factors constitute an appropriate benchmark for this data; the average R^2 for the multiple linear regressions is 0.899.

[18] We omit time series from the analysis when the fund is terminated during t+1 because otherwise obscure regression coefficients may skew the results.

	Quantile Portfolios										incr.	decr.
	1	2	3	4	5	6	7	8	9	10		
MDD:												
Return	0.0614	0.0627	0.0626	0.0655	0.0658	0.0660	0.0664	0.0673	0.0666	0.0588	x	x
α	−0.0011	−0.0065	−0.0074	−0.0069	−0.0073	−0.0072	−0.0068	−0.0051	−0.0029	−0.0056	x	x
β	0.0596	0.0670	0.0680	0.0688	0.0695	0.0697	0.0709	0.0723	0.0748	0.0787	✓	x
SMB	0.0001	0.0004	0.0008	0.0023	0.0024	0.0028	0.0034	0.0040	0.0045	0.0060	✓	x
HML	0.0018	0.0010	0.0007	0.0015	0.0023	0.0023	0.0018	0.0008	−0.0009	−0.0043	x	x
RMW	−0.0004	−0.0003	−0.0001	−0.0006	−0.0014	−0.0016	−0.0026	−0.0041	−0.0076	−0.0136	x	x
CMA	0.0014	0.0011	0.0006	0.0005	0.0003	0.0001	−0.0003	−0.0005	−0.0013	−0.0023	x	✓
ADD:												
Return	0.0669	0.0662	0.0648	0.0665	0.0639	0.0633	0.0637	0.0655	0.0640	0.0582	x	x
α	0.0018	−0.0061	−0.0066	−0.0064	−0.0077	−0.0060	−0.0064	−0.0042	−0.0049	−0.0103	x	x
β	0.0601	0.0671	0.0680	0.0690	0.0692	0.0700	0.0713	0.0727	0.0749	0.0769	✓	x
SMB	0.0015	0.0020	0.0021	0.0030	0.0018	0.0009	0.0024	0.0026	0.0041	0.0061	x	x
HML	0.0040	0.0036	0.0019	0.0019	0.0016	0.0003	−0.0006	−0.0009	−0.0020	−0.0028	x	x
RMW	−0.0018	−0.0017	−0.0010	−0.0009	−0.0011	−0.0016	−0.0026	−0.0041	−0.0070	−0.0105	x	x
CMA	0.0012	0.0014	0.0004	0.0000	0.0000	−0.0003	−0.0003	−0.0006	−0.0011	−0.0012	x	x
Std Dev.:												
Return	0.0640	0.0663	0.0652	0.0646	0.0632	0.0658	0.0651	0.0648	0.0636	0.0607	x	x
α	0.0008	−0.0058	−0.0093	−0.0101	−0.0094	−0.0073	−0.0062	−0.0050	−0.0037	−0.0009	x	x
β	0.0567	0.0657	0.0680	0.0695	0.0704	0.0711	0.0716	0.0727	0.0747	0.0790	✓	x
SMB	0.0009	0.0008	0.0011	0.0011	0.0008	0.0022	0.0028	0.0044	0.0059	0.0067	x	x
HML	0.0036	0.0030	0.0021	0.0015	0.0005	0.0010	0.0011	0.0000	−0.0016	−0.0041	x	x
RMW	0.0004	−0.0002	0.0008	0.0008	0.0005	−0.0012	−0.0040	−0.0054	−0.0088	−0.0153	x	x
CMA	0.0016	0.0028	0.0025	0.0017	0.0004	0.0000	−0.0002	−0.0019	−0.0029	−0.0045	x	x

Table 3.9: Average factor contributions to the mean excess return during t+1 in the Fama-French 5-factor model for funds sorted on MDD_{t-1}, ADD_{t-1}, or StdDev_{t-1}. Decile portfolio 1 contains the low drawdown funds.

by far compared to the other FF5 factors whose contributions are close to zero. Across decile portfolios, most of the factors' contributions remain fairly constant. Subtle trends across decile portfolios can be identified for the market factor and the robust minus weak operating profitability factor whose contributions slightly rise or fall, respectively, with increasing MDD_{t-1}, ADD_{t-1}, or StdDev_{t-1}. Broadly speaking, however, the factors' contributions to the return do not differ substantially between funds sorted on high and low drawdown. In summary, exploiting historical drawdown information about mutual funds sacrifices neither future returns nor future FF5 α, and does not alter the FF5 factors' contribution to the returns.
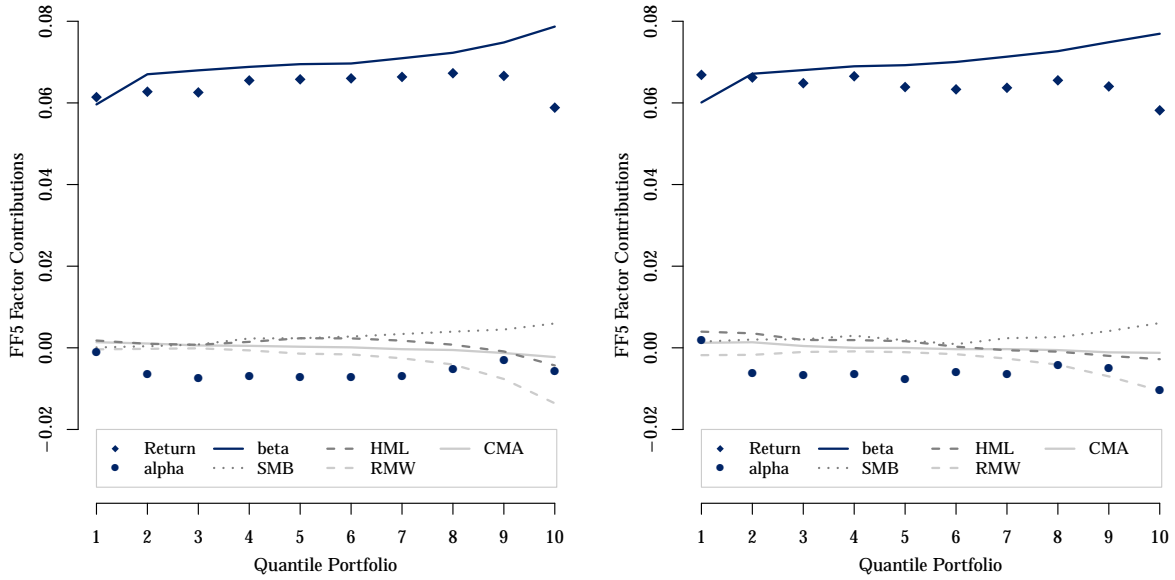
(a) Portfolios sorted according to MDD_{t-1}.

(b) Portfolios sorted according to ADD_{t-1}.
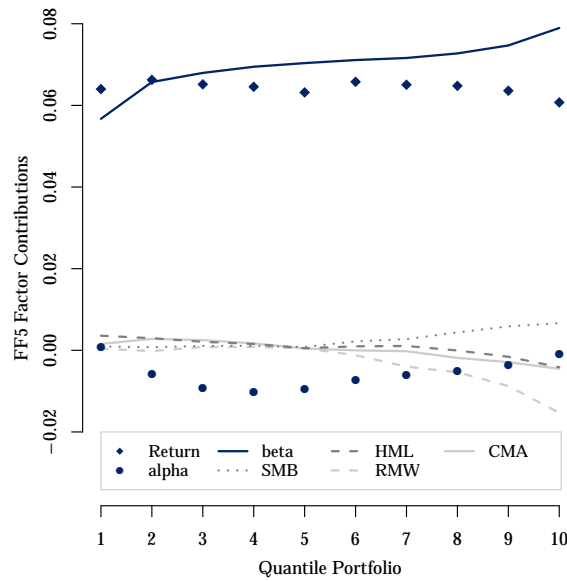
(c) Portfolios sorted according to StdDev_{t-1}.

Figure 3.4: Illustration of the average FF5 factor contributions to the excess return as provided in Table 3.9. For the 10 quantile portfolios, the diamonds indicate the excess return, the blue line indicates the market β, the dots refer to α, and the gray lines depict the contributions of the other FF5 factors (SMB, HML, RMW, CMA).

3.8 Conclusion

Choosing a mutual fund might be an individual's or institution's most far-reaching financial decision. It is likely to be informed by the fund's performance track record, both in terms of risk and returns. If these quantities are persistent, incorporating them into the investment decision makes immediate sense. However, if they are not persistent, reporting them offers no value in the first place. Although the literature on drawdown risk measures has been expanding, this important question of persistence has not yet been addressed. We fill this gap by analyzing the persistence of more than 7,000 mutual funds. Our principal finding is that the drawdown of mutual funds is highly persistent. For example, on average, funds in the highest ADD decile exhibit twice as much drawdown in the following year than funds in the lowest decile. Drawdown-based fund rankings between subsequent years are significantly positively correlated at approximately 0.40. Considerable differences in the average drawdown quantities between quantile portfolios underscore the economic significance of drawdown persistence. Not only is drawdown persistent for the full sample; similar results are observed for partitions with only U.S. funds, international funds, or sector funds. Further subdivision into fund categories (e.g., U.S. mid-cap growth funds) reveals that drawdown persists in all categories. Moreover, persistence holds for drawdown that is computed for three months, one year, and three years; it is also present irrespective of the period under investigation. Probing along different dimensions, we find ample and robust evidence of persistence.

The most obvious implication when drawdown persists is to predict future drawdown. However, is exploiting drawdown persistence the best option to infer future drawdown? Several alternatives are analyzed, and among them past standard deviation is found to be the best drawdown predictor. Although its outperformance (e.g., a rank correlation of 0.493 instead of 0.436) is not gargantuan, it is significantly present across all fund categories. As the standard deviation is often readily available, it may serve as a quick – and even more accurate – shortcut for estimating future drawdown. In conclusion, if you want to pick a fund with low future drawdown, the standard deviation should guide your choice. Further analysis of the mutual fund data and additional simulations help explain why these results occur. First, volatility persistence breeds drawdown persistence. Second, if returns were significantly persistent, the power of historical drawdown to predict future drawdown would increase strongly, while the power of historical standard deviation to predict future drawdown would not. Consequently, if the data at hand can be expected to exhibit significant return persistence, the drawdown measures themselves would constitute the superior indicator. Since strong return persistence cannot be expected in most

finance settings, using the shortcut will provide better results in most applications.

The additional analysis in Section 3.7 does not immediately relate to persistence. However, it addresses a pressing question when one decided to exploit drawdown persistence or the volatility shortcut. In such a case, an obvious question would be whether the choice to manage drawdown affects risk *and* return. We document that there is not a substantial effect on either returns or Fama-French 5-factor alphas. Therefore, there do not seem to be costs associated with reducing drawdown risks.

As this article is the first to investigate drawdown persistence, there is plenty of room for further research. Most obviously, the analysis may be expanded to include other drawdown measures, as we have restricted our analysis to two of the most common drawdown measures. Additionally, an analysis in conjunction with other types of risk measures may yield interesting results. The question of persistence, in which the same quantity is used during $t-1$ and $t+1$, is always closely linked to the question of predictability, in which other $t-1$ information may be exploited to improve the forecast of the quantity in $t+1$. By also considering other risk measures to predict drawdown, this article has moved toward improving drawdown prediction. However, much more extensive optimization is possible to improve how to choose funds with low future drawdown. This analysis focuses on mutual funds, which we deem a realistic use case. Usually, only a small number of funds are chosen by both institutional and private investors. Limiting risk and considering drawdown when making investment decisions can be advantageous, especially for important goals like retirement savings or college education funds. Nonetheless, our analysis may be replicated for other investment vehicles or stocks. As few people pick single stocks, examining portfolios of stocks may be of interest.

C Appendix

Panel A: Evaluation period: three months

Category Group	Persistence $\mathcal{P}_{\text{cor}}$	$\mathcal{P}_{10/1}$	Quantile Portfolios 1	2	3	4	5	6	7	8	9	10	incr.	# funds
MDD:														
All Equity	0.413	1.84	0.062	0.069	0.072	0.074	0.077	0.080	0.083	0.087	0.093	0.114	✓	7263
U.S. Equity	0.465	1.68	0.062	0.068	0.071	0.072	0.075	0.077	0.081	0.085	0.091	0.104	✓	4636
Int. Equity	0.381	1.64	0.066	0.073	0.075	0.077	0.079	0.080	0.083	0.088	0.094	0.108	✓	2064
Sector Equity	0.481	2.59	0.059	0.070	0.080	0.083	0.085	0.090	0.096	0.111	0.128	0.153	✓	781
ADD:														
All Equity	0.353	1.91	0.023	0.025	0.026	0.027	0.029	0.030	0.031	0.033	0.035	0.044	✓	7263
U.S. Equity	0.399	1.77	0.022	0.024	0.025	0.026	0.027	0.028	0.030	0.031	0.034	0.039	✓	4636
Int. Equity	0.327	1.68	0.025	0.028	0.029	0.029	0.030	0.031	0.032	0.034	0.037	0.042	✓	2064
Sector Equity	0.396	2.74	0.023	0.028	0.030	0.032	0.032	0.033	0.038	0.045	0.051	0.063	✓	781

Panel B: Evaluation period: three years

Category Group	Persistence $\mathcal{P}_{\text{cor}}$	$\mathcal{P}_{10/1}$	Quantile Portfolios 1	2	3	4	5	6	7	8	9	10	incr.	# funds
MDD:														
All Equity	0.448	1.62	0.216	0.230	0.234	0.240	0.248	0.259	0.269	0.281	0.298	0.350	✓	7263
U.S. Equity	0.445	1.49	0.211	0.223	0.227	0.230	0.237	0.245	0.256	0.269	0.283	0.314	✓	4636
Int. Equity	0.403	1.48	0.238	0.255	0.265	0.269	0.276	0.277	0.283	0.302	0.327	0.352	✓	2064
Sector Equity	0.463	1.98	0.219	0.237	0.256	0.266	0.276	0.294	0.312	0.339	0.374	0.434	✓	781
ADD:														
All Equity	0.422	1.91	0.066	0.069	0.072	0.075	0.079	0.084	0.087	0.091	0.098	0.126	✓	7263
U.S. Equity	0.417	1.63	0.062	0.066	0.067	0.070	0.072	0.076	0.081	0.083	0.088	0.101	✓	4636
Int. Equity	0.384	1.68	0.077	0.084	0.087	0.090	0.092	0.094	0.098	0.108	0.119	0.129	✓	2064
Sector Equity	0.392	2.58	0.071	0.084	0.092	0.092	0.090	0.093	0.101	0.119	0.146	0.183	x	781

Table 3.10: Drawdown persistence results for the full sample as in Table 3.1. Instead of one year, the evaluation period is **three months** (Panel A) and **three years** (Panel B).

Panel A: Persistence of MDD

Category	Persistence		Quantile Portfolios										incr.	# funds
	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10		
Large Value	0.407	1.43	0.056	0.062	0.065	0.066	0.068	0.069	0.070	0.072	0.074	0.080	✓	956
Large Blend	0.381	1.39	0.059	0.066	0.068	0.069	0.069	0.070	0.071	0.073	0.075	0.082	✓	1641
Large Growth	0.384	1.44	0.068	0.073	0.076	0.077	0.080	0.081	0.083	0.086	0.090	0.098	✓	1144
Mid-Cap Value	0.448	1.51	0.057	0.064	0.066	0.069	0.071	0.072	0.073	0.075	0.078	0.086	✓	456
Mid-Cap Blend	0.449	1.56	0.062	0.069	0.072	0.075	0.076	0.078	0.079	0.081	0.086	0.097	✓	647
Mid-Cap Growth	0.419	1.49	0.074	0.080	0.084	0.088	0.090	0.093	0.096	0.098	0.102	0.110	✓	723
Small Value	0.472	1.58	0.060	0.069	0.073	0.075	0.077	0.079	0.081	0.082	0.086	0.095	✓	408
Small Blend	0.374	1.41	0.073	0.078	0.080	0.083	0.086	0.087	0.088	0.089	0.091	0.103	✓	617
Small Growth	0.403	1.53	0.078	0.087	0.092	0.095	0.099	0.100	0.102	0.106	0.108	0.119	✓	596
Foreign Large Value	0.387	1.29	0.070	0.077	0.081	0.082	0.083	0.084	0.084	0.086	0.088	0.090	✓	206
Foreign Large Blend	0.304	1.25	0.069	0.074	0.076	0.077	0.078	0.079	0.079	0.081	0.082	0.086	✓	544
Foreign Large Growth	0.301	1.26	0.076	0.081	0.082	0.083	0.085	0.086	0.086	0.088	0.090	0.096	✓	245
Div. Emerging Mark.	0.338	1.29	0.084	0.093	0.098	0.100	0.101	0.102	0.102	0.104	0.105	0.108	✓	351
World Stock	0.350	1.45	0.060	0.067	0.069	0.071	0.073	0.075	0.078	0.080	0.081	0.087	✓	555
Real Estate	0.357	1.42	0.065	0.083	0.085	0.086	0.087	0.087	0.088	0.089	0.089	0.092	✓	133
Technology	0.427	1.52	0.095	0.110	0.115	0.118	0.119	0.120	0.124	0.126	0.133	0.144	✓	165

Panel B: Persistence of ADD

Category	Persistence		Quantile Portfolios										incr.	# funds
	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10		
Large Value	0.345	1.45	0.020	0.022	0.023	0.024	0.024	0.024	0.025	0.025	0.027	0.029	✓	956
Large Blend	0.325	1.43	0.021	0.023	0.024	0.024	0.024	0.025	0.025	0.026	0.027	0.030	✓	1641
Large Growth	0.294	1.44	0.025	0.027	0.028	0.028	0.029	0.030	0.030	0.031	0.032	0.036	✓	1144
Mid-Cap Value	0.379	1.60	0.020	0.023	0.024	0.024	0.025	0.026	0.026	0.027	0.029	0.032	✓	456
Mid-Cap Blend	0.384	1.64	0.022	0.025	0.026	0.027	0.028	0.028	0.029	0.030	0.031	0.036	✓	647
Mid-Cap Growth	0.349	1.56	0.027	0.030	0.031	0.032	0.033	0.035	0.035	0.037	0.038	0.042	✓	723
Small Value	0.381	1.64	0.022	0.025	0.027	0.028	0.029	0.029	0.030	0.031	0.032	0.036	✓	408
Small Blend	0.320	1.48	0.027	0.028	0.030	0.030	0.031	0.032	0.032	0.033	0.034	0.040	✓	617
Small Growth	0.348	1.55	0.029	0.032	0.034	0.036	0.037	0.038	0.039	0.040	0.041	0.045	✓	596
Foreign Large Value	0.324	1.30	0.027	0.030	0.031	0.031	0.031	0.031	0.032	0.033	0.034	0.035	✓	206
Foreign Large Blend	0.251	1.27	0.026	0.028	0.029	0.029	0.030	0.030	0.031	0.031	0.031	0.033	✓	544
Foreign Large Growth	0.231	1.23	0.030	0.031	0.031	0.032	0.033	0.034	0.033	0.033	0.034	0.037	x	245
Div. Emerging Mark.	0.263	1.30	0.033	0.037	0.039	0.039	0.040	0.041	0.041	0.041	0.042	0.043	✓	351
World Stock	0.271	1.48	0.023	0.025	0.026	0.027	0.028	0.028	0.029	0.030	0.030	0.034	✓	555
Real Estate	0.309	1.44	0.025	0.031	0.033	0.033	0.033	0.034	0.034	0.034	0.034	0.036	✓	133
Technology	0.358	1.47	0.038	0.043	0.045	0.045	0.047	0.048	0.049	0.050	0.053	0.056	✓	165

Table 3.11: Detailed drawdown persistence results for single fund categories as in Table 3.2. Instead of one year, the evaluation period is **three months**.

Panel A: Persistence of MDD

	Persistence		Quantile Portfolios											
Category	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	incr.	# funds
Large Value	0.313	1.21	0.195	0.207	0.212	0.217	0.218	0.219	0.222	0.225	0.226	0.236	✓	956
Large Blend	0.313	1.25	0.199	0.213	0.218	0.219	0.218	0.221	0.221	0.227	0.233	0.248	x	1641
Large Growth	0.360	1.30	0.224	0.236	0.239	0.241	0.248	0.253	0.256	0.264	0.274	0.291	✓	1144
Mid-Cap Value	0.333	1.25	0.207	0.222	0.226	0.230	0.232	0.236	0.240	0.244	0.243	0.258	x	456
Mid-Cap Blend	0.359	1.31	0.217	0.229	0.240	0.245	0.244	0.247	0.249	0.257	0.269	0.285	x	647
Mid-Cap Growth	0.380	1.32	0.251	0.264	0.272	0.284	0.291	0.299	0.300	0.307	0.312	0.331	✓	723
Small Value	0.421	1.38	0.216	0.238	0.245	0.249	0.254	0.260	0.263	0.263	0.278	0.298	✓	408
Small Blend	0.311	1.23	0.251	0.266	0.267	0.272	0.277	0.278	0.283	0.286	0.288	0.308	✓	617
Small Growth	0.373	1.32	0.271	0.291	0.304	0.310	0.321	0.327	0.327	0.334	0.341	0.357	✓	596
Foreign Large Value	0.329	1.15	0.263	0.272	0.282	0.286	0.284	0.295	0.292	0.297	0.307	0.303	x	206
Foreign Large Blend	0.262	1.16	0.244	0.261	0.269	0.267	0.267	0.268	0.275	0.272	0.279	0.283	x	544
Foreign Large Growth	0.286	1.17	0.287	0.305	0.306	0.306	0.308	0.311	0.311	0.310	0.318	0.336	x	245
Div. Emerging Mark.	0.293	1.19	0.327	0.359	0.368	0.370	0.372	0.373	0.379	0.386	0.377	0.389	x	351
World Stock	0.366	1.33	0.215	0.226	0.235	0.242	0.249	0.250	0.255	0.269	0.275	0.285	✓	555
Real Estate	0.188	1.17	0.246	0.279	0.276	0.273	0.273	0.266	0.274	0.275	0.279	0.288	x	133
Technology	0.389	1.35	0.311	0.356	0.364	0.375	0.374	0.374	0.384	0.385	0.391	0.419	x	165

Panel B: Persistence of ADD

	Persistence		Quantile Portfolios											
Category	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	incr.	# funds
Large Value	0.268	1.31	0.054	0.058	0.060	0.061	0.061	0.062	0.063	0.064	0.064	0.071	✓	956
Large Blend	0.301	1.34	0.059	0.063	0.065	0.065	0.065	0.066	0.068	0.070	0.072	0.079	✓	1641
Large Growth	0.326	1.45	0.071	0.076	0.077	0.080	0.081	0.084	0.086	0.089	0.092	0.103	✓	1144
Mid-Cap Value	0.312	1.37	0.054	0.061	0.060	0.060	0.063	0.065	0.065	0.066	0.069	0.074	x	456
Mid-Cap Blend	0.311	1.43	0.060	0.065	0.066	0.068	0.068	0.071	0.072	0.072	0.077	0.086	✓	647
Mid-Cap Growth	0.372	1.56	0.075	0.082	0.087	0.092	0.094	0.098	0.099	0.102	0.107	0.117	✓	723
Small Value	0.347	1.49	0.059	0.064	0.066	0.069	0.071	0.071	0.075	0.076	0.080	0.088	✓	408
Small Blend	0.254	1.39	0.072	0.076	0.078	0.076	0.078	0.079	0.081	0.083	0.084	0.100	x	617
Small Growth	0.365	1.61	0.082	0.092	0.099	0.103	0.104	0.107	0.110	0.114	0.117	0.132	✓	596
Foreign Large Value	0.299	1.24	0.083	0.086	0.090	0.092	0.094	0.093	0.096	0.095	0.099	0.103	x	206
Foreign Large Blend	0.249	1.28	0.079	0.088	0.089	0.090	0.089	0.090	0.092	0.092	0.094	0.101	x	544
Foreign Large Growth	0.222	1.17	0.099	0.103	0.105	0.108	0.109	0.112	0.113	0.111	0.115	0.116	x	245
Div. Emerging Mark.	0.276	1.25	0.116	0.126	0.132	0.136	0.136	0.137	0.140	0.141	0.142	0.145	✓	351
World Stock	0.335	1.56	0.064	0.071	0.075	0.077	0.080	0.078	0.082	0.089	0.089	0.100	x	555
Real Estate	0.231	1.27	0.075	0.086	0.089	0.091	0.089	0.091	0.089	0.090	0.093	0.095	x	133
Technology	0.357	1.41	0.117	0.134	0.141	0.145	0.145	0.147	0.151	0.152	0.162	0.165	✓	165

Table 3.12: Detailed drawdown persistence results for single fund categories as in Table 3.2. Instead of one year, the evaluation period is **three years**.

Panel A: Persistence of MDD

	Persistence		Quantile Portfolios											
Time Period	$\mathcal{P}_{\text{cor}}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	incr.	# funds
All Equity														
1990–1999	0.465	2.34	0.044	0.051	0.053	0.056	0.057	0.061	0.066	0.072	0.080	0.103	✓	2667
2000–2009	0.444	1.77	0.080	0.089	0.094	0.096	0.100	0.103	0.107	0.111	0.118	0.142	✓	5019
2010–2019	0.364	1.59	0.059	0.065	0.067	0.069	0.071	0.073	0.075	0.076	0.080	0.094	✓	4979
U.S. Equity														
1990–1999	0.494	2.14	0.043	0.050	0.053	0.054	0.056	0.059	0.063	0.069	0.078	0.092	✓	1811
2000–2009	0.466	1.62	0.081	0.088	0.092	0.094	0.098	0.101	0.104	0.109	0.116	0.131	✓	3388
2010–2019	0.451	1.48	0.058	0.063	0.065	0.066	0.069	0.071	0.073	0.076	0.078	0.086	✓	2992
Int. Equity														
1990–1999	0.383	2.00	0.053	0.055	0.056	0.058	0.060	0.060	0.061	0.072	0.082	0.106	✓	586
2000–2009	0.436	1.55	0.082	0.092	0.097	0.099	0.101	0.104	0.108	0.111	0.117	0.127	✓	1094
2010–2019	0.342	1.41	0.063	0.069	0.071	0.073	0.074	0.075	0.076	0.079	0.082	0.089	✓	1561
Sector Equity														
1990–1999	0.611	3.41	0.041	0.048	0.059	0.064	0.069	0.076	0.083	0.099	0.115	0.140	✓	266
2000–2009	0.490	2.38	0.076	0.091	0.105	0.108	0.108	0.114	0.124	0.141	0.161	0.181	✓	544
2010–2019	0.429	2.34	0.058	0.068	0.073	0.075	0.076	0.078	0.080	0.089	0.105	0.136	✓	543

Panel B: Persistence of ADD

	Persistence		Quantile Portfolios											
Time Period	$\mathcal{P}_{\text{cor}}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	incr.	# funds
All Equity														
1990–1999	0.375	2.60	0.015	0.017	0.018	0.019	0.020	0.021	0.023	0.025	0.029	0.039	✓	2667
2000–2009	0.364	1.78	0.032	0.035	0.037	0.038	0.040	0.041	0.043	0.045	0.047	0.057	✓	5019
2010–2019	0.335	1.67	0.021	0.022	0.023	0.024	0.025	0.026	0.027	0.027	0.029	0.035	✓	4979
U.S. Equity														
1990–1999	0.402	2.29	0.014	0.016	0.017	0.018	0.019	0.020	0.022	0.024	0.027	0.032	✓	1811
2000–2009	0.397	1.62	0.032	0.035	0.036	0.037	0.038	0.040	0.042	0.043	0.046	0.052	✓	3388
2010–2019	0.399	1.58	0.019	0.021	0.021	0.022	0.023	0.024	0.025	0.026	0.027	0.030	✓	2992
Int. Equity														
1990–1999	0.332	2.11	0.019	0.020	0.020	0.020	0.022	0.022	0.023	0.028	0.032	0.040	✓	586
2000–2009	0.325	1.50	0.034	0.037	0.039	0.040	0.041	0.042	0.044	0.044	0.046	0.051	✓	1094
2010–2019	0.327	1.59	0.022	0.025	0.026	0.027	0.028	0.028	0.029	0.030	0.032	0.035	✓	1561
Sector Equity														
1990–1999	0.513	3.93	0.015	0.018	0.021	0.024	0.026	0.027	0.031	0.038	0.046	0.059	✓	266
2000–2009	0.416	2.34	0.032	0.038	0.042	0.044	0.043	0.045	0.052	0.060	0.067	0.075	x	544
2010–2019	0.336	2.50	0.022	0.026	0.026	0.026	0.027	0.027	0.030	0.034	0.040	0.055	✓	543

Table 3.13: Subperiod analysis of the persistence results as in Table 3.3. Instead of one year, the evaluation period is **three months**.

Panel A: Persistence of MDD

Time Period	Persistence $\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	Quantile Portfolios 1	2	3	4	5	6	7	8	9	10	incr.	# funds
All Equity														
1990–1999	0.462	1.94	0.173	0.184	0.186	0.192	0.201	0.218	0.231	0.249	0.273	0.335	✓	2667
2000–2009	0.382	1.40	0.292	0.312	0.318	0.320	0.328	0.333	0.342	0.351	0.364	0.409	✓	5019
2010–2019	0.530	1.76	0.160	0.169	0.176	0.184	0.193	0.202	0.210	0.218	0.233	0.282	✓	4547
U.S. Equity														
1990–1999	0.473	1.79	0.164	0.172	0.179	0.183	0.190	0.198	0.216	0.233	0.256	0.293	✓	1811
2000–2009	0.359	1.30	0.293	0.310	0.310	0.312	0.318	0.325	0.330	0.342	0.353	0.380	✓	3388
2010–2019	0.557	1.61	0.152	0.162	0.168	0.171	0.179	0.187	0.198	0.207	0.217	0.244	✓	2758
Int. Equity														
1990–1999	0.420	1.81	0.203	0.209	0.225	0.229	0.237	0.237	0.247	0.287	0.327	0.368	✓	586
2000–2009	0.356	1.26	0.311	0.336	0.344	0.345	0.352	0.353	0.359	0.361	0.377	0.392	✓	1094
2010–2019	0.445	1.53	0.178	0.196	0.201	0.209	0.216	0.218	0.220	0.234	0.253	0.272	✓	1362
Sector Equity														
1990–1999	0.526	2.29	0.179	0.189	0.205	0.219	0.243	0.264	0.293	0.329	0.364	0.410	✓	266
2000–2009	0.434	1.70	0.281	0.311	0.343	0.347	0.343	0.363	0.381	0.405	0.432	0.479	x	544
2010–2019	0.473	2.23	0.180	0.189	0.195	0.208	0.219	0.233	0.233	0.255	0.301	0.401	✓	507

Panel B: Persistence of ADD

Time Period	Persistence $\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	Quantile Portfolios 1	2	3	4	5	6	7	8	9	10	incr.	# funds
All Equity														
1990–1999	0.400	2.28	0.050	0.051	0.051	0.054	0.058	0.064	0.069	0.073	0.081	0.114	✓	2667
2000–2009	0.341	1.54	0.101	0.109	0.114	0.117	0.121	0.125	0.127	0.129	0.134	0.156	✓	5019
2010–2019	0.544	2.82	0.034	0.036	0.039	0.041	0.045	0.049	0.053	0.059	0.068	0.096	✓	4547
U.S. Equity														
1990–1999	0.387	1.82	0.045	0.047	0.046	0.049	0.052	0.055	0.061	0.064	0.070	0.082	x	1811
2000–2009	0.344	1.46	0.099	0.107	0.109	0.112	0.115	0.119	0.124	0.125	0.131	0.145	✓	3388
2010–2019	0.543	2.14	0.029	0.031	0.032	0.034	0.036	0.039	0.042	0.045	0.048	0.062	✓	2758
Int. Equity														
1990–1999	0.415	2.11	0.064	0.065	0.066	0.068	0.070	0.073	0.081	0.103	0.122	0.135	✓	586
2000–2009	0.284	1.32	0.110	0.122	0.125	0.128	0.131	0.132	0.133	0.136	0.140	0.145	✓	1094
2010–2019	0.472	2.06	0.048	0.053	0.060	0.063	0.066	0.067	0.070	0.075	0.084	0.099	✓	1362
Sector Equity														
1990–1999	0.376	3.22	0.058	0.070	0.078	0.081	0.076	0.074	0.079	0.096	0.143	0.187	x	266
2000–2009	0.363	1.86	0.099	0.119	0.130	0.126	0.121	0.124	0.141	0.162	0.173	0.184	x	544
2010–2019	0.440	3.69	0.048	0.051	0.053	0.056	0.064	0.071	0.071	0.088	0.110	0.177	✓	507

Table 3.14: Subperiod analysis of the persistence results as in Table 3.3. Instead of one year, the evaluation period is **three years**.

	Persistence		Quantile Portfolios										
Category Group	$\mathcal{P}_{cor}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	incr.
MDD:													
All Equity	0.504	2.10	0.058	0.067	0.070	0.073	0.076	0.079	0.082	0.088	0.096	0.122	✓
	0.413	1.84	0.062	0.069	0.072	0.074	0.077	0.080	0.083	0.087	0.093	0.114	✓
U.S. Equity	0.560	1.85	0.059	0.066	0.069	0.072	0.074	0.077	0.081	0.086	0.094	0.109	✓
	0.465	1.68	0.062	0.068	0.071	0.072	0.075	0.077	0.081	0.085	0.091	0.104	✓
Int. Equity	0.460	1.79	0.062	0.071	0.074	0.077	0.079	0.081	0.084	0.088	0.096	0.111	✓
	0.381	1.64	0.066	0.073	0.075	0.077	0.079	0.080	0.083	0.088	0.094	0.108	✓
Sector Equity	0.609	3.02	0.053	0.062	0.072	0.079	0.085	0.092	0.101	0.117	0.135	0.160	✓
	0.481	2.59	0.059	0.070	0.080	0.083	0.085	0.090	0.096	0.111	0.128	0.153	✓
ADD:													
All Equity	0.438	2.33	0.021	0.025	0.026	0.027	0.028	0.029	0.031	0.033	0.036	0.049	✓
	0.353	1.91	0.023	0.025	0.026	0.027	0.029	0.030	0.031	0.033	0.035	0.044	✓
U.S. Equity	0.497	2.00	0.021	0.023	0.024	0.026	0.027	0.028	0.030	0.032	0.035	0.042	✓
	0.399	1.77	0.022	0.024	0.025	0.026	0.027	0.028	0.030	0.031	0.034	0.039	✓
Int. Equity	0.404	1.96	0.023	0.027	0.028	0.029	0.030	0.031	0.032	0.035	0.038	0.045	✓
	0.327	1.68	0.025	0.028	0.029	0.029	0.030	0.031	0.032	0.034	0.037	0.042	✓
Sector Equity	0.532	3.53	0.019	0.023	0.027	0.030	0.032	0.035	0.039	0.047	0.056	0.067	✓
	0.396	2.74	0.023	0.028	0.030	0.032	0.032	0.033	0.038	0.045	0.051	0.063	✓

Table 3.15: Analysis of how the standard deviation predicts drawdown as in Table 3.5. Instead of one year, the evaluation period is **three months**.

Category Group	Persistence		Quantile Portfolios										incr.
	$\mathcal{P}_{\text{cor}}$	$\mathcal{P}_{10/1}$	1	2	3	4	5	6	7	8	9	10	
MDD:													
All Equity	0.480	1.73	0.211	0.228	0.232	0.239	0.245	0.254	0.265	0.281	0.304	0.365	✓
	0.448	1.62	0.216	0.230	0.234	0.240	0.248	0.259	0.269	0.281	0.298	0.350	✓
U.S. Equity	0.529	1.62	0.204	0.217	0.221	0.226	0.234	0.243	0.254	0.273	0.293	0.331	✓
	0.445	1.49	0.211	0.223	0.227	0.230	0.237	0.245	0.256	0.269	0.283	0.314	✓
Int. Equity	0.405	1.47	0.237	0.252	0.261	0.271	0.279	0.281	0.287	0.302	0.326	0.349	✓
	0.403	1.48	0.238	0.255	0.265	0.269	0.276	0.277	0.283	0.302	0.327	0.352	✓
Sector Equity	0.600	2.26	0.199	0.209	0.238	0.255	0.272	0.291	0.313	0.362	0.416	0.449	✓
	0.463	1.98	0.219	0.237	0.256	0.266	0.276	0.294	0.312	0.339	0.374	0.434	✓
ADD:													
All Equity	0.444	2.28	0.061	0.067	0.069	0.073	0.076	0.080	0.084	0.092	0.104	0.139	✓
	0.422	1.91	0.066	0.069	0.072	0.075	0.079	0.084	0.087	0.091	0.098	0.126	✓
U.S. Equity	0.517	2.11	0.056	0.060	0.062	0.065	0.070	0.073	0.078	0.087	0.097	0.118	✓
	0.417	1.63	0.062	0.066	0.067	0.070	0.072	0.076	0.081	0.083	0.088	0.101	✓
Int. Equity	0.391	1.87	0.071	0.080	0.085	0.090	0.094	0.096	0.101	0.109	0.120	0.133	✓
	0.384	1.68	0.077	0.084	0.087	0.090	0.092	0.094	0.098	0.108	0.119	0.129	✓
Sector Equity	0.547	3.35	0.057	0.062	0.076	0.081	0.087	0.097	0.108	0.140	0.172	0.191	✓
	0.392	2.58	0.071	0.084	0.092	0.092	0.090	0.093	0.101	0.119	0.146	0.183	x

Table 3.16: Analysis of how the standard deviation predicts drawdown as in Table 3.5. Instead of one year, the evaluation period is **three years**.

Panel A: Correlations of relative ranks with MDD

	t−1			t+1		
t−1	Return	Std Dev.	MDD	Return	Std Dev.	MDD
Return	1	0.023	−0.350	0.078	0.069	0.010
Std Dev.		1	0.671	0.030	0.746	0.504
MDD			1	0.005	0.509	0.413

Panel B: Correlations of relative ranks with ADD

	t−1			t+1		
t−1	Return	Std Dev.	ADD	Return	Std Dev.	ADD
Return	1	0.023	−0.517	0.078	0.069	−0.050
Std Dev.		1	0.574	0.030	0.746	0.438
ADD			1	−0.024	0.419	0.353

Table 3.17: Correlations between different measures within the period or with the consecutive period as in Table 3.7. Instead of one year, the evaluation period is **three months**.

Panel A: Correlations of relative ranks with MDD

t−1	Return	Std Dev.	MDD	Return	Std Dev.	MDD
		t−1			t+1	
Return	1	−0.001	−0.358	0.137	0.087	0.026
Std Dev.		1	0.684	−0.064	0.626	0.480
MDD			1	−0.089	0.424	0.448

Panel B: Correlations of relative ranks with ADD

t−1	Return	Std Dev.	ADD	Return	Std Dev.	ADD
		t−1			t+1	
Return	1	−0.001	−0.512	0.137	0.087	−0.041
Std Dev.		1	0.600	−0.064	0.626	0.444
ADD			1	−0.114	0.364	0.422

Table 3.18: Correlations between different measures within the period or with the consecutive period as in Table 3.7. Instead of one year, the evaluation period is **three months**.

Panel A: Evaluation period: three months

Category Group	Persistence $\mathcal{P}_{cor}$	Quantile Portfolios 1	2	3	4	5	6	7	8	9	10	incr.	decr.
MDD:													
All Equity	0.005	0.020	0.020	0.020	0.020	0.020	0.019	0.019	0.019	0.017	0.011	x	x
U.S. Equity	0.000	0.022	0.022	0.021	0.021	0.022	0.021	0.021	0.020	0.019	0.016	x	x
Int. Equity	−0.023	0.019	0.017	0.016	0.015	0.013	0.013	0.011	0.010	0.010	0.009	x	x
Sector Equity	−0.037	0.025	0.024	0.020	0.022	0.023	0.025	0.023	0.015	0.003	−0.007	x	x
ADD:													
All Equity	−0.024	0.022	0.022	0.020	0.020	0.020	0.019	0.019	0.018	0.016	0.009	x	x
U.S. Equity	−0.028	0.024	0.023	0.022	0.022	0.021	0.021	0.021	0.020	0.019	0.015	x	x
Int. Equity	−0.046	0.020	0.017	0.016	0.015	0.014	0.013	0.012	0.010	0.010	0.007	x	✓
Sector Equity	−0.050	0.025	0.026	0.025	0.024	0.025	0.022	0.018	0.011	0.006	−0.008	x	x
Std Dev.:													
All Equity	0.030	0.019	0.018	0.019	0.019	0.018	0.018	0.020	0.020	0.020	0.014	x	x
U.S. Equity	0.026	0.021	0.021	0.021	0.020	0.020	0.021	0.021	0.021	0.021	0.020	x	x
Int. Equity	−0.013	0.019	0.016	0.015	0.014	0.013	0.013	0.013	0.012	0.010	0.009	x	x
Sector Equity	−0.018	0.024	0.022	0.018	0.021	0.023	0.023	0.020	0.013	0.007	0.002	x	x

Panel B: Evaluation period: three years

Category Group	Persistence $\mathcal{P}_{cor}$	Quantile Portfolios 1	2	3	4	5	6	7	8	9	10	incr.	decr.
MDD:													
All Equity	−0.089	0.238	0.234	0.235	0.233	0.222	0.219	0.216	0.217	0.217	0.166	x	x
U.S. Equity	−0.029	0.234	0.239	0.240	0.240	0.236	0.232	0.237	0.237	0.242	0.219	x	x
Int. Equity	−0.099	0.228	0.214	0.200	0.184	0.172	0.171	0.162	0.153	0.139	0.152	x	x
Sector Equity	−0.117	0.243	0.266	0.239	0.248	0.258	0.262	0.256	0.234	0.143	0.044	x	x
ADD:													
All Equity	−0.114	0.242	0.240	0.234	0.230	0.221	0.216	0.219	0.217	0.216	0.161	x	x
U.S. Equity	−0.042	0.239	0.241	0.239	0.236	0.230	0.230	0.232	0.242	0.242	0.226	x	x
Int. Equity	−0.123	0.234	0.212	0.201	0.188	0.177	0.165	0.167	0.152	0.142	0.138	x	x
Sector Equity	−0.130	0.246	0.254	0.233	0.244	0.260	0.268	0.269	0.223	0.133	0.064	x	x
Std Dev.:													
All Equity	−0.064	0.245	0.241	0.236	0.227	0.222	0.218	0.217	0.214	0.207	0.169	x	✓
U.S. Equity	−0.055	0.244	0.251	0.249	0.244	0.240	0.239	0.235	0.228	0.223	0.203	x	x
Int. Equity	−0.093	0.240	0.211	0.196	0.181	0.169	0.168	0.156	0.154	0.156	0.145	x	x
Sector Equity	−0.125	0.258	0.285	0.242	0.256	0.283	0.271	0.244	0.164	0.078	0.113	x	x

Table 3.19: Average returns for funds sorted on MDD$_{t-1}$, ADD$_{t-1}$, or StdDev$_{t-1}$ as in Table 3.8. Instead of one year, the evaluation horizon is **three months** (Panel A) and **three years** (Panel B).

Conclusion

The complexity of drawdown measures should not be underestimated. Both predicting their properties and making sense of results after an empirical analysis may be challenging. Drawdown properties with regard to return moments may serve as an example: Since drawdown measures are asymmetric and explicitly focused on losses, it is quite surprising that left-skewness does not significantly drive up all drawdown measures. No less surprising is that increased kurtosis leads to higher $CED_{0.95}$ but lower MDD. Even more, the finding that autocorrelation exerts diminishingly little influence on all drawdown measures – despite drawdown's explicit path dependence – is rather counter-intuitive. In other cases, plausible explanations can be found for at first glance startling findings: For example, the analysis within the third paper reveals that it is not the drawdown measure itself that predicts future drawdown best but the historical standard deviation. Further analysis helps understand this astonishing finding as a consequence of (lacking) return and (present) volatility persistence in the data. Similarly, the skill detection of drawdown performance ratios in the second paper seems peculiar at first, yet it can be explained as a consequence of unreliable performance ratios in periods of negative returns. Overall, intuitions about drawdown should be verified adequately instead of jumping to conclusion too quickly.

Without repeating the results from each chapters' conclusion, what can be learned about drawdown? In contrast to the standard deviation, drawdown is highly sensitive to expected returns. Since risk measures are usually associated only with variability and uncertainty, this relation undermines the clear distinction between risk and return, which is implicitly used in ratios such as the Sharpe ratio. For drawdown measures, both numerator and denominator of such risk–return ratios are sensitive to the return. As detailed in the second paper, this amplifies problems inherent in all performance ratios. Since performance ratios assume a rather peculiar structure in general[19], and all drawdown ratios are found to be fairly similar, the use of drawdown

---

[19]It takes a small difference (0.095) in risk but a much larger difference (0.95) in return to get a realistic risk–return ratio of $\frac{0.05}{0.1} = 0.5$ up to a ratio of 10.

performance ratios is questionable in principle.

Understanding the influence of autocorrelation on drawdown measures is non-trivial. Autocorrelation on short horizons is found to exert no effect on drawdown measures; autocorrelation on long horizons on the other hand is found to strongly influence drawdown persistence. Crucially, the assessment of autocorrelation requires controlling for volatility because both are closely intertwined, and – without adequate controls – possible results differ drastically. Although autocorrelation is assessed using different models, all results remain dependent on the choice of the simulation model. In order to thoroughly grasp the relationship between autocorrelation and drawdown, more research is needed, where models, time horizons, and control variables are varied systematically.

Some applications of drawdown measures are more promising than others. Due to their path dependence, their focus on downside risk and the intuitive reference point, drawdown measures can theoretically improve how risk is assessed. Either by themselves or in addition to standard risk measures, they can potentially help quantify what human beings truly perceive as risk. Consequentially, the application of drawdown measures is suitable for money management, where anticipating and managing risk perceptions is vital. For example, funds could state that they have experienced particularly little drawdown historically, and investing with them would, thus, have induced little pain. Drawdown could also be used by fund managers internally by incorporating drawdown information into fund management decisions to maximize client satisfaction and minimize the likelihood of sudden withdrawals or fund closure. Managers could devise strategies to minimize the probability of breaching a drawdown constraint or include drawdown measures into the objective function when optimizing portfolios. Apart from intuitive risk assessment, are drawdown measures a means to improve investment decisions along other dimensions? So far, there is little evidence that historical drawdown might be useful indicator either to improve other risk measures, market timing, or asset allocation. However, a lack of predictive power for other quantities does not invalidate the use of drawdown measures in other applications.

Despite adhering to the same fundamental structure, quite different risk measures belong to the class of drawdown measures. These differences reflect in highly different weighting schemes in the wDD framework, different moment behavior, different skill detection, and different ranking of investments. Considering the entirety of results, what drawdown measure could be recommended for applications in practice? According to the rank correlation results in the second paper, ADD, lwDD, and ADD^2 are highly correlated. Of these particularly robust measures, which all

incorporate the entire drawdown graph, the ADD is arguably the most intuitive and established measure. It may be complemented by either MDD or eopDD with which it is rather weakly correlated. The CED cannot be recommended because it is not applicable whenever ex-post evaluation is required. Moreover, like the CDD with any confidence level α other than zero or one, the CED always needs justification why a particular α is used. As it is rarely a good idea to put all one's eggs into one basket, utilizing several drawdown measures to grasp such an important and diverse concept as risk can be recommended. Moreover, the wDD framework provides practitioners with an accessible tool to devise a drawdown risk measure that suits their personal preferences.

As drawdown measures are underrepresented in the literature, this dissertation can only begin to shed light on the mechanics of drawdown risk measures. Hence, the potential for future research is abundant. As discussed already, numerous and conflicting definitions of drawdown measures exist in the literature. For instance, the definition of Schuhmacher and Eling (2011) based on absolute losses derived from monthly uncompounded cumulative excess returns bears almost no similarity to the drawdown defined in this dissertation. One might assume that differences in definition are limited to special drawdown measures, but the opposite is true: even for the most standard maximum drawdown, definitions from different sources potentially result in different sizes and locations of the MDD. Therefore, it might be interesting to assess empirically how much of a difference the choice of a drawdown definition actually makes for a given asset path. Arguably, some definitions are objectively better than others, for example, because they are more robust. Hence, it may be worthwhile to sift through all definitions in the literature, weight their strengths and weaknesses, and compile a list of well-defined and useful drawdown measures.

Not only for drawdown definitions but also for different drawdown measures, theoretical properties might be worth a systematic assessment. Usually, when new drawdown measures are defined in the literature, some properties such as homogeneity, subadditivity, or convexity are tested. However, the sets of properties and their definitions vary, making results from different sources hard to compare. Moreover, some properties are more relevant for drawdown measures than others; for example, some of the coherence axioms of Artzner et al. (1999), such as monotonicity, are rather meaningless for drawdown measures, while some axioms of generalized deviation measures of Rockafellar et al. (2006), such as shift invariance, are arguably more appropriate. It might be worthwhile to first discuss which properties are most desirable for drawdown measures, then test these properties for a broad set of drawdown measures using identical definitions, and compare the results for different measures systematically.

When drawdown statistics for several investments, say mutual funds, are compared in practice, the absolute size of the drawdown strongly depends on the market phase. Most mutual fund drawdowns tend to be large during a downturn, and small in a bull market. To eliminate this market effect, Bradford and Siliski (2016) define the so-called active drawdown measure where each asset's drawdown is calculated with respect to a suitable benchmark index. Such a benchmark-relative measure might be fairer to fund managers because their individual influence on the market phase is negligible. The computation of the active drawdown is fairly straightforward because the drawdown graph is simply computed not from the asset price but instead from the difference between cumulative returns of the asset and its benchmark. As Bradford and Siliski (2016) only provide an unnecessarily laborious explanation how to compute the active drawdown, its properties are still unexplored. Potential new research might, for example, investigate the persistence of the maximum active drawdown.

Regarding behavioral finance, further research is certainly needed to substantiate the common claim that drawdown measures capture risk more intuitively than other risk measures. Although there are many plausible reasons why an asymmetric, loss-focused, reference-based drawdown measure should capture risk perceptions of human investors more accurately than other risk measures, thorough empirical evidence has not yet been collected. A first attempt in this direction is the study by Harris et al. (2015), who analyze crowd-sourced data of preferences between two asset paths. They observe that the drawdown ratios' choice of the more desirable path particularly often coincides with the human choice. Many other experimental setups are conceivable: Study participants could be asked to rank more than two asset paths according to risk or "attractiveness". More actively, study participants could trade in an artificial stock market where they see track records, drawdown figures, or both, and their decisions to buy or sell a stock could be correlated with implicitly present or explicitly stated drawdown properties. In such a setting, traders could also be asked to indicate their satisfaction with past decisions, which could be expected to deteriorate in situations of high drawdown. Variations of the experimental setup could help understand the different psychological components of drawdown and address which drawdown measures best capture risk preferences and perceptions.

Instead of altogether new approaches, parts of the three papers could, of course, be modified to extend or replicate their findings. As all setups were chosen deliberately, these following modifications are – from the author's perspective – usually slightly less suitable. In the first paper, an analysis of the relation between a stock's true moments and its drawdown becomes possible by simulations. The same assessment with real-world data would have to be based

on correlations between *realized* moments and *realized* drawdown, which might be interesting nonetheless. When it comes to assessing stock picking skill in the second paper, no adequate alternative seems available to replace the simulation model in which skill can be added via the hit ratio. Rank correlations between different investments could, however, be computed for alternative data sets. Yet, both too extreme differences between assets (e.g., a bond fund vs. an equity fund) and too little differences between assets (e.g., two index funds with slightly different weighting schemes) are unlikely to yield relevant comparisons between risk measures. When it comes to persistence in the third paper, utilizing different data might be interesting because the presence of drawdown persistence in other asset classes is still an open question. In order to assess persistence via simulations, different processes or at least different parametrizations would have to be chosen to mirror the heterogeneity of funds. In other words, since persistence is measured relative to a cross section, the simulations would have to create some meaningful cross section. When drawdown measures are deemed relevant in practice, strategies become necessary on how to choose assets or construct portfolios such that drawdown quantities are met. Although the assessment of persistence is a step in this direction, more research could be directed at finding an efficient drawdown frontier, or identifying strategies to limit future drawdown in different asset classes. In the third paper, a quick analysis is performed on the relation between drawdown and return. It may be extended to a thorough analysis of a potential low risk anomaly with drawdown measures, i.e., low drawdown risk coinciding with high returns. Since several low risk anomalies have been documented for other risk measures (e.g., in Schneider et al. (2020)), it is not entirely unlikely that some form of low drawdown anomaly exists.

If this dissertation had to be condensed into three sentences, it might be these: Various intuitive and versatile drawdown risk measures can be defined in the wDD framework. Most of them are able to detect stock picking skill, and they are highly persistent for mutual funds, but their use in drawdown performance ratios is problematic. Interesting and sometimes counter-intuitive results occur when analyzing drawdown measures and higher return moments, autocorrelation, and the prediction of future drawdown.

Bibliography

Aas, K., I. H. Haff, and X. K. Dimakos (2005). Risk Estimation Using the Multivariate Normal Inverse Gaussian Distribution. *Journal of Risk 5*(2), 39–60.

Adcock, C., M. Eling, and N. Loperfido (2015). Skewed Distributions in Finance and Actuarial Science: a Review. *European Journal of Finance 21*(13-14), 1253–1281.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data* (2nd ed.). New Jersey: John Wiley & Sons.

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

Alexander, G. J. and A. M. Baptista (2006). Portfolio Selection With a Drawdown Constraint. *Journal of Banking & Finance 30*(11), 3171–3189.

Alexander, G. J. and A. M. Baptista (2017). Bank Capital Regulation of Trading Portfolios: An Assessment of the Basel Framework. *Journal of Money, Credit and Banking 49*(4), 603–634.

Ammann, M., O. Huber, and M. Schmid (2013). Hedge Fund Characteristics and Performance Persistence. *European Financial Management 19*(2), 209–250.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The Cross-Section of Volatility and Expected Returns. *Journal of Finance 61*(1), 259–299.

Angoshtari, B., E. Bayraktar, and V. Young (2016). Optimal Investment to Minimize the Probability of Drawdown. *Stochastics 88*(6), 946–958.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent Measures of Risk. *Mathematical Finance 9*(3), 203–228.

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.

Aye, G. C., M. Balcilar, R. Gupta, N. Kilimani, A. Nakumuryango, and S. Redford (2014). Predicting BRICS Stock Returns using ARFIMA Models. *Applied Financial Economics 24*(17), 1159–1166.

Bacon, C. (2008). *Practical Portfolio Performance Measurement and Attribution* (2nd ed.). Chichester: John Wiley & Sons.

Bai, L. and P. Liu (2019). Drawdown and Drawup for Fractional Brownian Motion with Trend. *Journal of Theoretical Probability 32*(3), 1581–1612.

Baillie, R. T. (1996). Long Memory Processes and Fractional Integration in Econometrics. *Journal of Econometrics 73*(1), 5–59.

Baker, M., B. Bradley, and J. Wurgler (2011). Benchmarks as Limits to Arbitrage: Understanding the Low Volatility Anomaly. *Financial Analysts Journal 67*(1), 40–54.

Ball, C. A. and W. N. Torous (1983). A Simplified Jump Process for Common Stock Returns. *Journal of Financial and Quantitative Analysis 18*(1), 53–65.

Barberis, N., M. Huang, and T. Santos (2001). Prospect Theory and Asset Prices. *The Quarterly Journal of Economics 66*(1), 1–50.

Barndorff-Nielsen, O. E. (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics 24*, 1–13.

Barndorff-Nielsen, O. E. (1998). Processes of Normal Inverse Gaussian Type. *Finance and Stochastics 2*, 41–68.

Barndorff-Nielsen, O. E., T. Mikosch, and S. I. Resnick (2013). *Lévy Processes: Theory and Applications.* Boston: Birkhäuser.

Barndorff-Nielsen, O. E. and R. Stelzer (2005). Absolute Moments of Generalized Hyperbolic Distributions and Approximate Scaling of Normal Inverse Gaussian Lévy Processes. *Scandinavian Journal of Statistics 32*(4), 617–637.

Barras, L., O. Scaillet, and R. Wermers (2010). False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas. *Journal of Finance 65*(1), 179–216.

Baxter, G. and M. D. Donsker (1957). On the Distribution of the Supremum Functional for Processes with Stationary Independent Increments. *Transactions of the American Mathematical Society 85*, 73–87.

Bell, D. E. (1982). Regret in Decision Making under Uncertainty. *Operations Research 30*(5), 961–981.

Bell, D. E. (1983). Risk Premiums for Decision Regret. *Management Science 29*(10), 1156–1166.

Benartzi, S. and R. Thaler (1995). Myopic Loss Aversion and the Equity Premium Puzzle. *The Quarterly Journal of Economics 110*(1), 73–92.

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

Bhardwaj, G. and N. R. Swanson (2006). An Empirical Investigation of the Usefulness of ARFIMA Models for Predicting Macroeconomic and Financial Time Series. *Journal of Econometrics 131*(1–2), 539–578.

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.

Bleichrodt, H., A. Cillo, and E. Diecidue (2010). A Quantitative Measurement of Regret Theory. *Management Science 56*(1), 161–175.

Bleichrodt, H. and P. P. Wakker (2015). Regret Theory: A Bold Alternative to the Alternatives. *Economic Journal 125*(583), 493–532.

Bollen, N. P. and J. A. Busse (2005). Short-Term Persistence in Mutual Fund Performance. *Review of Financial Studies 18*(2), 569–597.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T. (1987). On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process. *Journal of Time Series Analysis 9*(2), 121–131.

Bradford, D. and D. Siliski (2016). Performance Drawdowns in Asset Management: Extending Drawdown Analysis to Active Returns. *Journal of Performance Measurement 21*(1).

Brown, S. and W. N. Goetzmann (1995). Performance Persistence. *Journal of Finance 50*(2), 679–698.

Burghardt, G. and B. Walls (2011). *Managed Futures for Institutional Investors*. New Jersey: John Wiley & Sons.

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

Busse, J. A., A. Goyal, and S. Wahal (2010). Performance and Persistence in Institutional Investment Management. *Journal of Finance 65*(2), 765–790.

Camara Leal, R. P. and B. V. de Melo Mendes (2005). Maximum Drawdown: Models and Applications. *Journal of Alternative Investments 7*(4), 83–91.

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.

Caporale, G. M. and L. A. Gil-Alana (2006). Modelling Stochastic Volatility in Asset Returns Using Fractionally Integrated Semiparametric Techniques. *Applied Financial Economics Letters 2*(1), 9–12.

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*(2011-10), 1–34.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance 52*(1), 57–82.

Casarin, R., M. Lazzarin, L. Pelizzon, and D. Sartore (2005). Relative Benchmark Rating and Persistence Analysis: Evidence from Italian Equity Funds. *European Journal of Finance 11*(4), 297–308.

Casati, A. and S. Tabachnik (2013). The Statistics of the Maximum Drawdown in Financial Time Series. In J. A. Batten, P. MacKay, and N. Wagner (Eds.), *Advances in Financial Risk Management*. London: Palgrave Macmillan.

Challet, D. (2017). Sharper Asset Ranking from Total Drawdown Durations. *Applied Mathematical Finance 24*(1), 1–22.

Chan, K., K. C. Chan, and G. A. Karolyi (1991). Intraday Volatility in the Stock Index and Stock Index Futures Markets. *Review of Financial Studies 4*(4), 657–684.

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.

Cheng, I.-H., H. Hong, and J. A. Scheinkman (2015). Yesterday's Heroes: Compensation and Risk at Financial Firms. *Journal of Finance 70*(2), 839–879.

Cheridito, P., A. Nikeghbali, and E. Platen (2012). Processes of Class Sigma, Last Passage Times, and Drawdowns. *SIAM Journal on Financial Mathematics 3*(1), 280–303.

Cherny, V. and J. Obłój (2013). Portfolio Optimisation under Non-Linear Drawdown Constraints in a Semimartingale Financial Model. *Finance and Stochastics 17*(4), 771–800.

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

Cohen, R. B., J. D. Coval, and L. Pástor (2005). Judging Fund Managers by the Company they Keep. *Journal of Finance 60*(3), 1057–1096.

Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes.* London: Chapmann & Hall / CRC Press.

Cvitanic, J. and I. Karatzas (1995). On Portfolio Optimization under 'Drawdown' Constraints. *IMA Lecture Notes in Mathematics & Applications 65*, 77–88.

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.

Defusco, R. A., G. V. Karels, and K. Muralidhar (1996). Skewness Persistence in US Common Stock Returns: Results from Bootstrapping Tests. *Journal of Business Finance & Accounting 23*(8), 1183–1195.

Dennis, J. E. and R. B. Schnabel (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Philadelphia: Society for Industrial and Applied Mathematics.

Diecidue, E. and J. Somasundaram (2017). Regret Theory: A new Foundation. *Journal of Economic Theory 172*, 88–119.

Ding, Z. and C. W. Granger (1996). Modeling Volatility Persistence of Speculative Returns: A New Approach. *Journal of Econometrics 73*(1), 185–215.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance 1*(1), 83–106.

Dittmar, R. F. (2002). Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns. *Journal of Finance 57*(1), 369–403.

Doney, R. A. (2007). Spectrally Negative Lévy Processes. In J. Picard (Ed.), *Fluctuation Theory for Levy Processes.* Berlin, Heidelberg: Springer.

Douady, R., A. N. Shiryaev, and M. Yor (2000). On Probability Characteristics of 'Downfalls' in a Standard Brownian Motion. *Theory of Probability and its Applications 44*(1), 29–38.

El Ghoul, S. and A. Karoui (2017). Does Corporate Social Responsibility Affect Mutual Fund Performance and Flows? *Journal of Banking & Finance 77*, 53–63.

Elie, R. and N. Touzi (2008). Optimal Lifetime Consumption and Investment Under a Drawdown Constraint. *Finance and Stochastics 12*(3), 299–330.

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

Eling, M. (2009). Does Hedge Fund Performance Persist? Overview and New Empirical Evidence. *European Financial Management 15*(2), 362–401.

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.

Elton, E. J., M. J. Gruber, and C. R. Blake (1996). The Persistence of Risk-Adjusted Mutual Fund Performance. *Journal of Business 69*(2), 133–157.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*(4), 987–1007.

Ergün, A. T. (2011). Skewness and Kurtosis Persistence: Conventional vs. Robust Measures. *SSRN Electronic Journal.*

Eriksson, A., E. Ghysels, and F. Wang (2009). The Normal Inverse Gaussian Distribution and the Pricing of Derivatives. *The Journal of Derivatives 16*(3), 23–37.

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.

Fama, E. F. and K. R. French (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics 116*(1), 1–22.

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

Frazzini, A. and L. H. Pedersen (2014). Betting Against Beta. *Journal of Financial Economics 111*(1), 1–25.

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.

Gilli, M. and E. Schumann (2009). An Empirical Analysis of Alternative Portfolio Selection Criteria. *Swiss Finance Institute Research Paper No. 09-06, University of Geneva*.

Goddard, J., H. Liu, P. Molyneux, and J. O. Wilson (2011). The Persistence of Bank Profit. *Journal of Banking & Finance 35*(11), 2881–2890.

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

Gollier, C. (2020). Aversion to Risk of Regret and Preference for Positively Skewed Risks. *Economic Theory 70*(4), 913–941.

Granger, C. W. (1980). Long Memory Relationships and the Aggregation of Dynamic Models. *Journal of Econometrics 14*(2), 227–238.

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

Grinblatt, M. and S. Titman (1992). The Persistence of Mutual Fund Performance. *Journal of Finance 47*(5), 1977–1984.

Grossman, S. and Z. Zhou (1993). Optimal Investment Strategies for Controlling Drawdowns. *Mathematical Finance 3*(3), 241–276.

Guidolin, M. and A. Timmermann (2008). International Asset Allocation under Regime Switching, Skew, and Kurtosis Preferences. *Review of Financial Studies 21*(2), 889–935.

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*(1), 61–72.

Hadjiliadis, O. and J. Vecer (2006). Drawdowns Preceding Rallies in the Brownian Motion Model. *Quantitative Finance 6*(5), 403–409.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hanson, F. B. (2007). *Applied Stochastic Processes and Control for Jump-Diffusions*. Philadelphia: Society for Industrial and Applied Mathematics.

Harri, A. and B. W. Brorsen (2004). Performance Persistence and the Source of Returns for Hedge Funds. *Applied Financial Economics 14*(2), 131–141.

Harris, G., A. Panangadan, and V. K. Prasanna (2015). Learning of Performance Measures from Crowd-Sourced Data with Application to Ranking of Investments. In T. Cao, E. P. Lim, Z. H. Zhou, T. B. Ho, D. Cheung, and H. Motoda (Eds.), *Advances in Knowledge Discovery and Data Mining*. Cham: Springer.

Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller (2010). Portfolio Selection with Higher Moments. *Quantitative Finance 10*(5), 469–485.

Harvey, C. R. and Y. Liu (2018). Detecting Repeatable Performance. *Review of Financial Studies 31*(7), 2499–2552.

Harvey, C. R. and A. Siddique (2000). Conditional Skewness in Asset Pricing Tests. *Journal of Finance 55*(3), 1263–1295.

Haugen, R. A. and A. J. Heins (1975). Risk and the Rate of Return on Financial Assets: Some Old Wine in New Bottles. *Journal of Financial and Quantitative Analysis 10*(5), 775–784.

Heidorn, T., D. Kaiser, and C. Roder (2009). The Risk of Funds of Hedge Funds: An Empirical Analysis of the Maximum Drawdown. *Journal of Wealth Management 12*(2), 89–100.

Hendricks, D., J. Patel, and R. Zeckhauser (1993). Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974–1988. *Journal of Finance 48*(1), 93–130.

Homm, U. and C. Pigorsch (2012). Beyond the Sharpe Ratio: An Application of the Aumann-Serrano Index to Performance Measurement. *Journal of Banking and Finance 36*(8), 2274–2284.

Hosking, J. R. M. (1981). Fractional Differencing. *Biometrika 68*(1), 165–176.

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.

Hui, K. W., K. K. Nelson, and P. E. Yeung (2016). On the Persistence and Pricing of Industry-Wide and Firm-Specific Earnings, Cash Flows, and Accruals. *Journal of Accounting and Economics 61*(1), 185–202.

Huij, J. and M. Verbeek (2007). Cross-Sectional Learning and Short-Run Persistence in Mutual Fund Performance. *Journal of Banking & Finance 31*(3), 973–997.

Investment Company Institute (2019). *2019 Investment Company Fact Book* (59th ed.). Washington D.C.: Investment Company Institute.

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

Jacobsen, B. and D. Dannenburg (2003). Volatility Clustering in Monthly Stock Returns. *Journal of Empirical Finance 10*(4), 479–503.

Jegadeesh, N. and S. Titman (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance 48*(1), 65–91.

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.

Jondeau, E. and M. Rockinger (2003). Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements. *Journal of Economic Dynamics and Control 27*(10), 1699–1737.

Jondeau, E. and M. Rockinger (2012). On the Importance of Time Variability in Higher Moments for Asset Allocation. *Journal of Financial Econometrics 10*(1), 84–123.

Kahneman, D., J. Knetsch, and R. Thaler (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of the Political Economy 98*(6), 1325–1348.

Kahneman, D. and M. W. Riepe (1998). Aspects of Investor Psychology. *Journal of Portfolio Management 24*(4), 52–65.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(3), 263–291.

Kalemanova, A., B. Schmid, and R. Werner (2007). The Normal Inverse Gaussian Distribution for Synthetic CDO Pricing. *The Journal of Derivatives 14*(3), 80–94.

Kardaras, C., J. Obłój, and E. Platen (2017). The Numéraire Property and Long-Term Growth Optimality for Drawdown-Constrained Investments. *Mathematical Finance 27*(1), 68–95.

Kendall, M. G. (1945). The Treatment of Ties in Ranking Problems. *Biometrika 33*(3), 239–251.

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

Kim, S.-H. (2018). Are Value at Risk and Maximum Drawdown Different from Volatility in Stock Market? *Journal of Applied Business Research 34*(2), 217–223.

Korn, O., P. M. Möller, and C. Schwehm (2019). Drawdown Measures: Are They All the Same? *Journal of Portfolio Management, forthcoming*.

Kosowski, R., A. Timmermann, R. Wermers, and H. White (2006). Can Mutual Fund 'Stars' Really Pick Stocks? New Evidence from a Bootstrap Analysis. *Journal of Finance 61*(6), 2551–2595.

Kou, S. G. (2002). A Jump-Diffusion Model for Option Pricing. *Management Science 48*(8), 1086–1101.

Kwásnicki, M., J. Małecki, and M. Ryznar (2013). Suprema of Lévy Processes. *Annals of Probability 41*(3 B), 2047–2065.

Landriault, D., B. Li, and S. Li (2015). Analysis of a Drawdown-Based Regime-Switching Lévy Insurance Model. *Insurance: Mathematics and Economics 60*, 98–107.

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 (2017a). A Unified Approach for Drawdown (Drawup) of Time-Homogeneous Markov Processes. *Journal of Applied Probability 54*(2), 603–626.

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

Lemmon, M. L., M. R. Roberts, and J. F. Zender (2008). Back to the Beginning: Persistence and the Cross-Section of Corporate Capital Structure. *Journal of Finance 63*(4), 1575–1608.

Lhabitant, F. S. (2004). *Hedge Funds: Quantitative Insights*. Chichester: John Wiley & Sons.

Lin, C. H., W. H. Huang, and M. Zeelenberg (2006). Multiple Reference Points in Investor Regret. *Journal of Economic Psychology 27*(6), 781–792.

Ma, L., Y. Tang, and J. P. Gómez (2019). Portfolio Manager Compensation in the U.S. Mutual Fund Industry. *Journal of Finance 74*(2), 587–638.

Magdon-Ismail, M., A. F. Atiya, A. Pratap, and Y. S. Abu-Mostafa (2004). On the Maximum Drawdown of a Brownian Motion. *Journal of Applied Probability 41*(1), 147–161.

Mahmoud, O. (2017). The Temporal Dimension of Risk. *Journal of Risk 19*(3), 57–83.

Malkiel, B. G. (1995). Returns from Investing in Equity Mutual Funds 1971 to 1991. *Journal of Finance 50*(2), 549–572.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business 36*(4), 394–419.

Maringer, D. and P. Parpas (2009). Global Optimization of Higher Order Moments in Portfolio Selection. *Journal of Global Optimization 43*(2-3), 219–230.

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

Mateus, I. B., C. Mateus, and N. Todorovic (2019). Use of Active Peer Benchmarks in Assessing UK Mutual Fund Performance and Performance Persistence. *European Journal of Finance 25*(12), 1077–1098.

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

Merton, R. C. (1976). Option Pricing when Underlying Stock Returns are Discontinuous. *Journal of Financial Economics 3*, 125–144.

Mijatović, A. and M. R. Pistorius (2012). On the Drawdown of Completely Asymmetric Lévy Processes. *Stochastic Processes and their Applications 122*(11), 3812–3836.

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

Molyboga, M. and C. L'Ahelec (2016). Portfolio Management with Drawdown-Based Measures. *Journal of Alternative Investments 19*(3), 75–89.

Muralidhar, K. (1993). The Bootstrap Approach for Testing Skewness Persistence. *Management Science 39*(4), 487–491.

Nanda, V. K., Z. J. Wang, and L. Zheng (2009). The ABCs of Mutual Funds: On the Introduction of Multiple Share Classes. *Journal of Financial Intermediation 18*(3), 329–361.

Nath, R. (1996). A Note on Testing for Skewness Persistence. *Management Science 42*(1), 138–141.

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

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

Pistorius, M. R. (2004). On Exit and Ergodicity of the Spectrally One-sided Lévy Process Reflected at its Infimum. *Journal of Theoretical Probability 17*(1), 183–220.

Pivetta, F. and R. Reis (2007). The Persistence of Inflation in the United States. *Journal of Economic Dynamics and Control 31*(4), 1326–1358.

Pospisil, L. and J. Vecer (2010). Portfolio Sensitivity to Changes in the Maximum and the Maximum Drawdown. *Quantitative Finance 10*(6), 617–627.

Rieder, U. and M. Wittlinger (2014). On Optimal Terminal Wealth Problems with Random Trading Times and Drawdown Constraints. *Advances in Applied Probability 46*(1), 121–138.

Roche, H. (2019). Asset Management with Endogenous Withdrawals under a Drawdown Constraint. *Quantitative Finance 19*(2), 289–312.

Rockafellar, R. T., S. Uryasev, and M. Zabarankin (2006). Generalized Deviations in Risk Analysis. *Finance and Stochastics 10*(1), 51–74.

Rydberg, T. H. (1997). The Normal Inverse Gaussian Lévy Process: Simulation and Approximation. *Communications in Statistics. Part C: Stochastic Models 13*(4), 887–910.

Schneider, P., C. Wagner, and J. Zechner (2020). Low-Risk Anomalies? *Journal of Finance 75*(5), 2673–2718.

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.

Scott, R. C. and P. A. Horvath (1980). On the Direction of Preference for Moments of Higher Order than the Variance. *Journal of Finance 35*(4), 915–919.

Sekine, J. (2013). Long-Term Optimal Investment with a Generalized Drawdown Constraint. *SIAM Journal on Financial Mathematics 4*(1), 452–473.

Sénéchal, E. (2010). The Empirical Law of Active Management: Perspectives on the Declining Skill of U.S. Fund Managers. *Journal of Portfolio Management 37*(1), 121–132.

Sharpe, W. F. (1966). Mutual Fund Performance. *Journal of Business 39*(1), 119–138.

Sharpe, W. F. (1975). Adjusting for Risk in Portfolio Performance Measurement. *Journal of Portfolio Management 1*(2), 29–34.

Sharpe, W. F. (1994). The Sharpe Ratio. *Journal of Portfolio Management 21*(1), 49–58.

Shefrin, H. and M. Statman (1985). The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *Journal of Finance 40*(3), 777–790.

Shefrin, H. M. and M. Statman (1984). Explaining Investor Preference for Cash Dividends. *Journal of Financial Economics 13*, 253–282.

Shi, Y., X. Cui, J. Yao, and D. Li (2015). Dynamic Trading with Reference Point Adaptation and Loss Aversion. *Operations Research 63*(4), 789–806.

Singleton, J. C. and J. Wingender (1986). Skewness Persistence in Common Stock Returns. *Journal of Financial and Quantitative Analysis 21*(3), 335–341.

Strahilevitz, M. A., T. Odean, and B. M. Barber (2011). Once Burned, Twice Shy: How Naive Learning, Counterfactuals, and Regret Affect the Repurchase of Stocks Previously Sold. *Journal of Marketing Research 48*, 102–120.

Sun, Q. and Y. Yan (2003). Skewness Persistence with Optimal Portfolio Selection. *Journal of Banking & Finance 27*(6), 1111–1121.

Tashman, A. and R. J. Frey (2009). Modeling Risk in Arbitrage Strategies using Finite Mixtures. *Quantitative Finance 9*(5), 495–503.

Unser, M. (2000). Lower Partial Moments as Measures of Perceived Risk: An Experimental Study. *Journal of Economic Psychology 21*(3), 253–280.

van Hemert, O., M. Ganz, C. R. Harvey, S. Rattray, E. S. Martin, and D. Yawitch (2020). Drawdowns. *SSRN Electronic Journal*.

Vecer, J. (2006). Maximum Drawdown and Directional Trading. *Risk 19*(12), 88–92.

Vecer, J. (2007). Preventing Portfolio Losses by Hedging Maximum Drawdown. *Wilmott 5*(4), 1–8.

Veld, C. and Y. V. Veld-Merkoulova (2008). The Risk Perceptions of Individual Investors. *Journal of Economic Psychology 29*(2), 226–252.

Wilhelmsson, A. (2009). Value at Risk with Time Varying Variance, Skewness and Kurtosis - the NIG-ACD Model. *Econometrics Journal 12*(1), 82–104.

Wu, Y. X., S. L. Shen, and D. J. Yuan (2016). Characteristics of Dewatering Induced Drawdown Curve Under Blocking Effect of Retaining Wall in Aquifer. *Journal of Hydrology 539*, 554–566.

Yao, H., Y. Lai, Q. Ma, and H. Zheng (2013). Characterization of Efficient Frontier for Mean-Variance Model with a Drawdown Constraint. *Applied Mathematics and Computation 220*, 770–782.

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

Zabarankin, M., K. Pavlikov, and S. Uryasev (2014). Capital Asset Pricing Model (CAPM) with Drawdown Measure. *European Journal of Operational Research 234*(2), 508–517.

Statement of contribution
to each paper of the cumulative dissertation

To the three papers of the cumulative dissertation I personally contributed as follows:

1. to the paper *Drawdown Measures and Return Moments* (single-author paper): conceptualization, empirical analysis, and writing, in total: 100%,

2. to the paper *Drawdown Measures: Are They All the Same?* co-authored by Olaf Korn and Christian Schwehm: significant contributions to the conceptualization, empirical analysis, and writing, in total: 70%, and

3. to the paper *Drawdown Persistence and a Convenient Shortcut to Predicting Mutual Fund Drawdown* (single-author paper): conceptualization, empirical analysis, and writing, in total: 100%.

Göttingen, 18.11.2020

_____   _____

Place, Date   Philipp M. Möller