Drawdowns

OTTO VAN HEMERT, MARK GANZ, CAMPBELL R. HARVEY, SANDY RATTRAY, EVA SANCHEZ MARTIN, and DARREL YAWITCH[1]

This version: May 4th, 2020

ABSTRACT

Common risk metrics reported in academia include volatility, skewness, and factor exposures. The maximum drawdown statistic is rarely calculated, perhaps because it is path dependent and estimated with greater uncertainty. In practice, however, asset managers and fiduciaries routinely use the drawdown statistic for fund allocation and redemption decisions. To help such decisions, we begin by quantifying the probability of hitting a certain drawdown level, given various return distribution properties. Next, we show that drawdown-based rules can be particularly useful for improving investment performance over time by detecting managers that lose their ability to outperform. This can happen as a result of structural market changes, increased competition for the type of strategy employed, staff turnover or a fund accumulating too many assets. Finally, we show that drawdown-based rules can be used as a risk reduction technique, but this impacts both expected returns and risk.

Keywords: Trading strategies, alpha, outperformance, crowding, downside risk, skewness, hitting time, allocation, redemption, Type I error, Type II error, drawdown, Sharpe ratio, structural breaks, Corona crash, COVID-19 crash.

[1] Harvey is at Duke University, the National Bureau of Economic Research and is an adviser to Man Group plc. The other authors were at Man Group plc during this research. The authors would like to thank Nick Barberis, Anthony Ledford, Matthew Sargaison, and Rikiya Takahashi for valuable comments. Please direct correspondence to: Otto.VanHemert@man.com.

1. Introduction

When evaluating managers or strategies, investors pay close attention to the maximum drawdown, i.e., the largest peak-to-trough return over the life of an investment. For example, for hedge fund investments, money is often pulled out when a threshold for the maximum drawdown is crossed. The maximum drawdown statistic is appealing, as it is unambiguous in its calculation and captures the most unfavorable investment outcome: buying at the peak and selling at the bottom.

The maximum drawdown statistic is different from other metrics such as volatility and downside measures like skewness or semi-variance in that it crucially depends on the *order* in which the returns occur. Closed-form solutions are hard to obtain, except under very restrictive assumptions.[2] In the first part of this paper, we conduct a simulation study to determine the sensitivity of the probability of reaching a given maximum drawdown threshold to key assumptions. We call these the "drawdown Greeks". Key drivers of the maximum drawdown that we identify are: the evaluation horizon (time to dig a hole), Sharpe ratio (ability to climb out of a hole), and the persistence in risk (chance of having a losing streak). The latter may motivate a manager to actively target a more stable risk profile over time when facing strict drawdown limits. We find that non-normal, but still time-independent, returns – for example the occasional gap move down – only matter much when they are large compared to what we generally observe for a range of financial markets. The reason is that with independent returns, the central limit theorem kicks in: multi-period returns start to look more normal as one increases the number of periods.

Next, we compare the ability of different manager replacement rules to improve investment performance over time. We introduce a framework to decide whether to replace a manager (or strategy). This decision will be subject to two types of errors: a Type I error of replacing a good manager and a Type II error of mistakenly not firing a bad manager.[3] We also recognize that the timing of these replacements matters, as a bad manager can do more harm the longer they are managing assets.

When managers are of constant (but unknown) quality, a replacement rule based on the total return is typically preferred. Clearly, as the total return makes full use of all historical return data available. However, drawdown-based rules are more suitable when there is a meaningful chance that managers lose their skill over time. In practice, this can happen as a result of structural market changes, staff turnover, increased competition for the type of strategy employed (crowding), or a fund accumulating too much assets.

Reducing the allocation to an underperforming manager using drawdown-based rules can be seen as a halfway house between no action and immediate replacement of the manager. However, if such risk reductions are not compensated for by increasing risk elsewhere in the portfolio, they will generally lead to lower expected returns. That is, unless a manager's conditional expected

[2] See for example Magdon-Ismail et al. (2004), where they study the behavior of maximum drawdown for the case of a Brownian motion with drift and an analytic expression is derived for the expected value of maximum drawdown (with zero drift) and infinite series representation (for nonzero drift).
[3] See also Harvey and Liu (2020) for an analysis of the tradeoff between Type I and Type II errors, as well as their differential costs.

returns (in excess of the cash rate) turn negative. This requires one to believe a manager is actually value destroying (in which case immediate replacement seems a more appropriate step) or if one believes there is very high degree of persistence in returns and previous returns were negative.

Finally, we summarize the main results and discuss five key takeaways for allocators choosing among managers, or for managers choosing between different investments strategies.

We have not tried to identify the impact of drawdown rules on manager behavior. But we are very aware that the presence of a "drawdown rule" will itself cause managers to act differently – nobody likes getting fired. From this perspective, a drawdown rule might be considered to have some similarities to volatility scaling; managers who show behavioral aversion to being fired will reduce risk near down the drawdown limit. From this perspective, drawdown rules might be considered 'poor man's volatility scaling'.[4]

2. Drawdown Greeks

In this section, we explore how sensitive the likelihood of hitting a certain drawdown level is to key drivers, like the Sharpe ratio, evaluation time window, and autocorrelation of returns. Borrowing terminology from option pricing theory, we call these sensitivities the "drawdown Greeks".[5]

2.1 Probability distribution for maximum drawdown level

We start with a simple setting of normal, independent, and identically distributed (IID) monthly returns.[6] In Figure 1, we show the probability distribution of the maximum drawdown statistic for our baseline case: 10 year time window, 10% annualized volatility, 0.5 annualized Sharpe ratio. Throughout we rely on simulation, where each parameterization is evaluated with 100,000 simulations of monthly returns for the evaluation window.

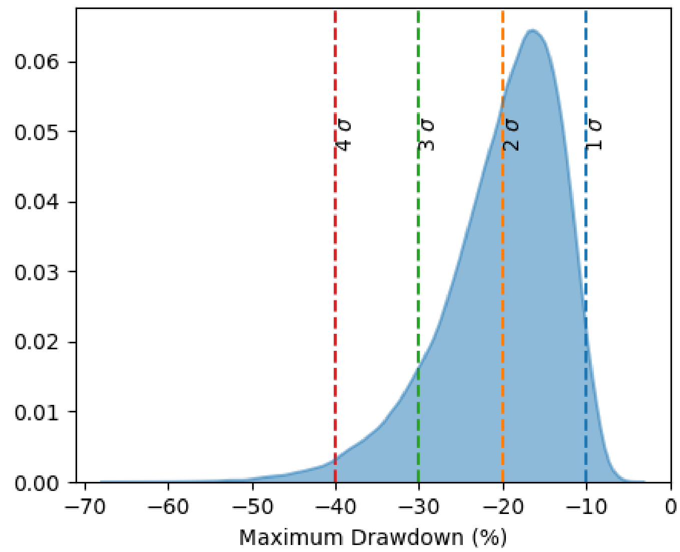
We highlight with vertical lines maximum drawdown levels of 1, 2, 3, and 4 annual standard deviation (or sigma) moves, corresponding to -10%, -20%, -30%, and -40% drawdown levels. The associated probability of reaching a maximum drawdown of that level of worse is given by the area under the curve to the left of the associated vertical line. It is 97.1%, 43.0%, 9.9%, and 1.5% for 1,

[4] See Harvey et al. (2018) for a discussion on the impact of volatility scaling on risk and return characteristics.
[5] This section adds to a vast literature on drawdowns that includes: 1) statistical characteristics, see, e.g., Douady, Shiryaev, and Yor (2000), Magdon-Ismail et al. (2004), Hadjiliadis and Vecer (2006), Casati and Tabachnik (2012), Bailey and Lopez de Prado (2015), and Busseti et al. (2016); 2) portfolio optimization, see, e.g., Grossman and Zhou (1993), Chekhlov, Uryasev and Zabarankin (2005), Cvitanic et al. (2019); 3) hedging and risk management, see, e.g., Carr, Zhang and Hadjiliadis (2011), Leal and Mendes (2005); 4) trading strategies; see, e.g., Vecer (2006); 5) measurement, see, e.g., Korn, Möller and Schwehm (2020); and 6) economic mechanisms, see, e.g., Sornette (2003).
[6] Our analysis is based on monthly, rather than daily, return data for two reasons. First, we think investment and allocation decisions by large institutions are more likely to take place at a monthly frequency. Second, returns at the daily frequency are harder to model as they are influenced by a pronounced intra-month variation in the news flow; e.g., bigger moves on the day major economic news is released. Monthly returns are somewhat better behaved, as they reflect the combination of both high- and low-news days. The more complicated case of daily drawdown evaluation and replacement decisions is left for future research.

2, 3, and 4 sigma levels, respectively. So in almost half of the cases, one reaches a drawdown of two full annual standard deviations (or -20%) over the 10-year period, even though the annual Sharpe ratio is a respectable 0.5. One-in-ten cases, one even reaches a drawdown of three full annual standard deviations (or -30%).

Figure 1: Probability distribution for the maximum drawdown statistic

The figure shows probability distribution for the maximum drawdown statistic using normal, IID monthly returns over a 10-year window with a 10% annualized volatility and 0.5 annualized Sharpe ratio (the baseline case). The vertical, dashed lines correspond to drawdowns of size 1, 2, 3, and 4 annual standard deviations.



2.2 Drawdown Greeks without higher order effects

Next, we consider how deviations from the baseline assumptions impact the probability of hitting a drawdown level. In Figure 2, we illustrate how the probability of a given level of maximum drawdown changes if we modify one of the following assumptions at a time: (A) annualized volatility, 10% baseline; (B) time window, 10 years baseline; (C) annualized Sharpe ratio, 0.5 baseline; and (D) autocorrelation, 0.0 baseline.[7]

In Panel A, we show how the probability of a given maximum drawdown changes when we vary the standard deviation of the return process, while holding constant the Sharpe ratio. The orange line represents the probability of a maximum drawdown that is -2 sigma (annual return standard deviations), or worse.[8] This value is 43% for the baseline case (see also the discussion of Figure 1), indicated by the vertical dashed line. The orange line is near horizontal, i.e., varying the standard deviation of returns hardly changes the probability of reaching a certain maximum drawdown level, as long as you assume the Sharpe ratio stays constant and the threshold is expressed in terms of sigmas. That is, the probability of a 20% maximum drawdown when returns have a 10% standard deviation is similar to the probability of 10% maximum drawdown when returns have a 5% standard deviation (assuming the Sharpe ratio is 0.5 in both cases; i.e., the expected returns

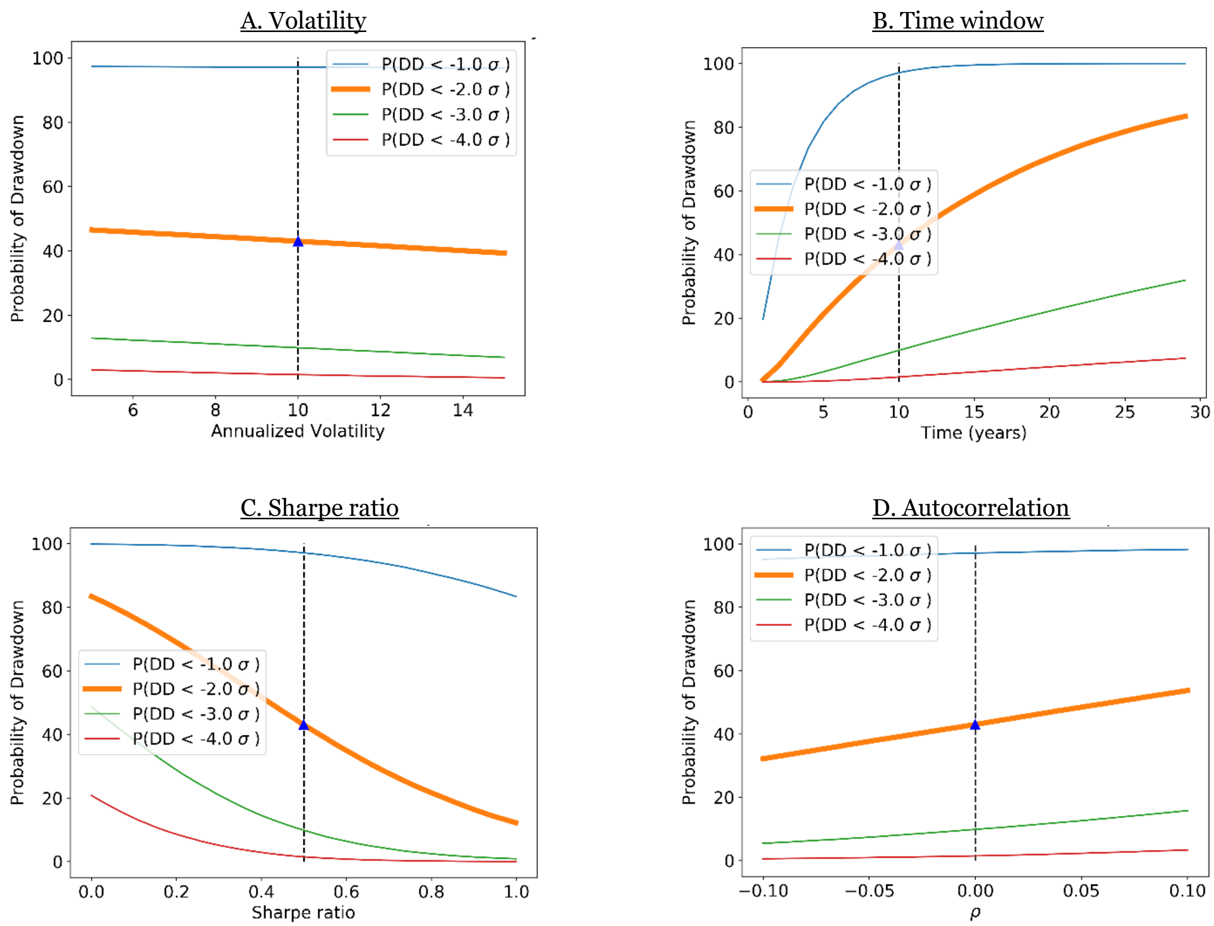
[7] Bailey and Lopez de Prado (2015) argue that ignoring the effect of serial correlation in the return generating process leads to a gross underestimation of the downside potential of hedge fund strategies.

[8] Both the variable we vary on the horizontal axis and the -2 sigma cutoff are based on the (ex-ante) standard deviation for the return process. Probabilities (vertical axis) are based on average realized values.

increase as volatility increases). The orange line is near horizontal, but not exactly. In fact, it is gently sloping downwards, reflecting the influence of compounding of returns.

Figure 2: Sensitivity of the probability of a maximum drawdown to key parameters

The figure shows probability of reaching a maximum drawdown of 1, 2, 3, and 4 sigma (annual standard deviations). In all cases returns are normal and identically distributed. In the different panels, we vary the (A) volatility holding constant the Sharpe ratio, (B) time window, (C) Sharpe ratio, and (D) autocorrelation. The vertical dashed line corresponds to the baseline case of 10% annualized volatility, 10-year window, 0.5 annualized Sharpe ratio, and 0.0 autocorrelation.



In Panel B, we illustrate the impact of changing the evaluation time horizon. The baseline case is 10 years. As a return stream is evaluated over a longer window, the probability of hitting a certain drawdown level naturally increases.

In Panel C, we vary the Sharpe ratio, while holding the constant standard deviation of returns. In the default case, we have an annualized Sharpe ratio of 0.5. The impact of Sharpe ratio on the probability of reaching a certain maximum drawdown level is large, which is intuitive, as the Sharpe ratio captures the ability to lift yourself out of a hole. It is exactly this effect that investors using drawdown rules are hoping to isolate – the low Sharpe ratio managers will be removed by the presence of the rule.

In Panel D, we vary the correlation, ρ, between time t and time $t-1$ monthly returns. In the formula below, the $\tilde{\mu}$ and $\tilde{\sigma}$ terms capture the unconditional mean and standard deviation, respectively, where we use a tilde to make clear it concerns monthly returns (in contrast to, e.g., Figure 2, where we used σ, without tilde, for the annualized standard deviation). The mean and standard deviations are pre-multiplied with a term featuring ρ to offset the effect of non-zero autocorrelation on the mean and standard deviation:

$$R_{t+1} = (1-\rho)\tilde{\mu} + \rho R_t + \sqrt{(1-\rho^2)}\tilde{\sigma}\varepsilon_{t+1},$$

where ε is standard normal and independent and identically distributed (IID).

We illustrate the impact of autocorrelation in monthly returns for values ranging from -0.1 to +0.1. We consider an autocorrelation of 0.1 (or similarly -0.1) a large value, as it implies a large degree of predictability.[9] The impact of a 0.1 autocorrelation in monthly returns on the expected maximum drawdown (versus a baseline value of 0) is comparable to that of reducing the Sharpe ratio from 0.5 to 0.4.

2.3 Bootstrapped US equity returns

Next, we bootstrap 2-year blocks from US equity returns since 1926 with monthly returns scaled to have 10% unconditional volatility.[10] Using actual return realizations allows us to determine if the inference is different from our simulated, normally distributed returns. Selecting blocks, rather than individual months, is to preserve the original time-series structure within a block

In the left panel of Figure 3, we present the sensitivity to the time window, holding the Sharpe ratio constant at 0.5 (by adjusting the mean returns appropriately). In the right panel of Figure 3, we present the sensitivity to the Sharpe ratio, while holding the time window constant at 10 years. As such, these figures can be directly compared to Panels B and C in Figure 2, where we simulated from a normal, independent, and identical distribution.

For the case of a 10-year window and Sharpe ratio of 0.5, we had a 43% probability of hitting a 2-sigma drawdown (see Figure 2, baseline case for our simulated returns). This probability increases to 55% for the bootstrapped actual returns in Figure 3 (see dashed line crossing orange line). This increased probability of hitting a drawdown level is a result of both non-normality of monthly returns and heteroscedasticity (clustering of volatility). This is illustrated in Appendix A for US equity returns since 1926, with volatility persistently high around, for example, 1929 (the Great Depression) and 2008 (the Global Financial Crisis).

---

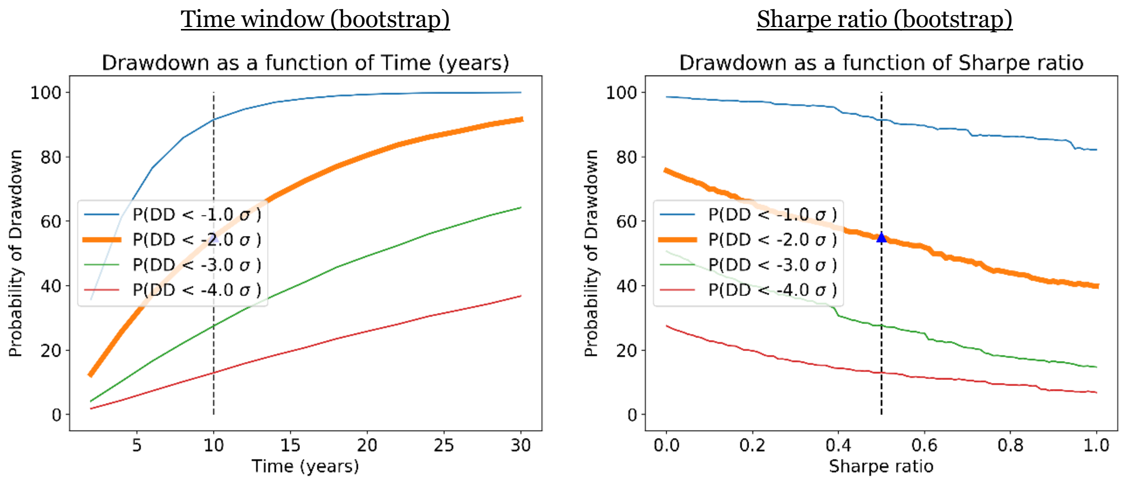
[9] To motivate this statement, consider the monthly process defined above and set $\tilde{\mu} = 0$ for simplicity. If $R_t = +\tilde{\sigma}$ that means the annualized conditional Sharpe ratio $\frac{E_t[R_{t+1}]}{\sigma_t[R_{t+1}]}\sqrt{12} = \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{12}$, which equals a value of 0.35 for a monthly autocorrelation of ρ=0.1. And if $R_t = +2\tilde{\sigma}$, the annualized conditional Sharpe ratio is 0.70. These are high values compared to a typical annualized Sharpe ratio for general equities (say 8%/20% = 0.4) or a typical hedge fund (with Sharpe ratio of, say 0.5 to 1.0). Note that ex-post one can measure a higher realized autocorrelation for a given market; our statement is just that on an ex-ante basis, a monthly autocorrelation of 0.1 already provides a high degree of predictability.

[10] This is an approach where one randomly selects (with replacement) blocks of consecutive observations from an actual distribution. As a robustness check, we reran our analysis with blocks longer than 24 months, and found similar results. See Efron and Tibshirani (1986) for an early discussion of bootstrap methods.

While non-normality and volatility clustering tends to increase the probability to hit a certain drawdown level, the sensitivity to the time window and Sharpe ratio looks very similar between the normal (Figure 2) and non-normal, bootstrapped case (Figure 3).

Figure 3: Drawdown probability with bootstrapped US equity returns

The figure shows probability of reaching a maximum drawdown of 1, 2, 3, and 4 sigma (annual standard deviations), as a function of the time window (left panel) and Sharpe ratio (right panel). We bootstrap 2-year blocks from US equity returns since 1926. We hold constant the unconditional volatility at 10%. In the left panel we hold constant the Sharpe ratio at 0.5. In the right panel we hold constant the time window at 10 years.



2.4 The impact of gap risk

Financial markets can experience sudden, negative returns of a magnitude that is implausible under the assumption of normally distributed returns. An example at the time of writing is the "Corona Crash" of March 2020. That is, markets can experience a gap move down. To illustrate this point, in Appendix B, we list for a range of securities the worst negative monthly return, expressed as a number of (annualized) standard deviations (last column). We see that for the 25 to 50 years of available history, the worst monthly returns are -1 to -1.5 annual standard deviations (which corresponds to 3.5 to 5.2 monthly standard deviations).[11]

We will explore the impact on the expected maximum drawdown of having a monthly move equal to $-k$ annual standard deviations with a 1% probability (once every 8.3 years on average). We adjust the mean of the returns in the other 99% cases, so that the average return is held constant while we vary the size of the gap move. In other words, we have the following distribution, where we continue to assume returns are IID:

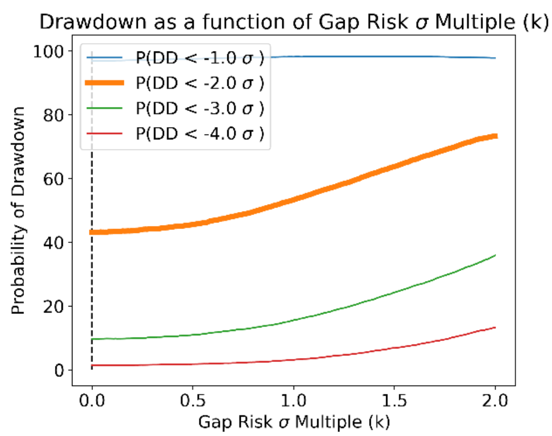
$$R = baseline\ case + \begin{cases} -k\sigma & with\ 1\%\ probability \\ +\dfrac{1}{99}k\sigma & with\ 99\%\ probability. \end{cases}$$

[11] Under a normal distribution, a -5 sigma or worse monthly move happens less than once every 250,000 years.

In Figure 4, we vary the size of the gap move (k in the formula above). The baseline corresponds to $k = 0$. Note that $k = 1$ is already a large value. Appendix Table A1 shows that the largest monthly return is just greater than 1, but this is over typically a 25- to 50-year window, rather than a 10-year period. While the probability of a large drawdown indeed increases with an increased probability of a gap move, the impact is somewhat limited for a $k = 1$ (monthly move equal to 1 annual standard deviation, or 3.5 monthly standard deviations). We think this is intuitive, as such a move doesn't immediately take you through a, say, -2 sigma, drawdown limit. Additionally, the drivers of returns in the medium to long term, like the Sharpe ratio, remain the key drivers of the probability to hit a drawdown.

Figure 4: Impact of gap risk

The figure shows probability of reaching a maximum drawdown of 1, 2, 3, and 4 sigma (annual standard deviations). In the baseline case (indicated by the vertical dashed line), we have normal, independent and identically distributed (IID) monthly returns, with an annualized mean, standard deviation, and Sharpe ratio of 5%, 10%, and 0.5, respectively, and an evaluation window of 10 years. We vary the size of the gap move (occurring in 1% of the months, or once every 8.3 years on average), with a zero gap size corresponding to the baseline case.



3. Manager replacement rules

Investors face considerable uncertainty around the quality of the managers (or strategies) when selecting them. Moreover, after investment, a manager's quality may deteriorate for a variety of reasons, including crowding of the investment style, excessive asset gathering by the manager, or a less favorable macroeconomic backdrop. This raises the question how to deal with a situation like the one illustrated in Figure 5, where the total return for a manager still looks quite healthy, but the recent drawdown looks worrisome. Did something change? Was the manager never good in the first place?

To navigate the uncertainty around the quality of managers, investors need a framework for deciding whether to replace a manager or not. Otherwise, behavioral biases can lead to sub-optimal decisions, as e.g. illustrated by Goyal and Wahal (2008), who show that investors are too quick to hire and to fire managers.[12]

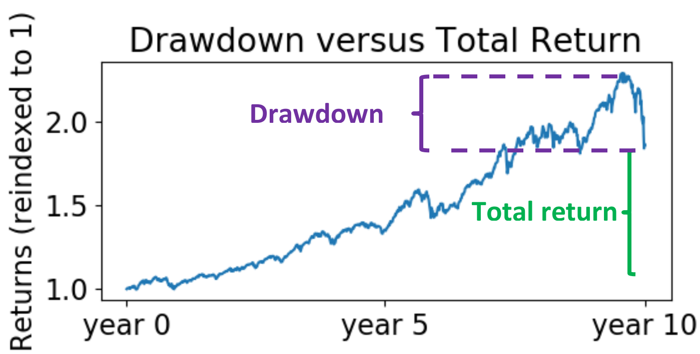
[12] Behavioral biases may arise because drawdowns can be attention grabbing when observed in a graph like Figure 5. Such an effect has its foundation in the salience theory, see Bordalo, Gennaioli, and Shleifer (2012). There is also a notion that after experiencing a painful loss, people become more sensitive to any additional losses – they just can't

To illustrate the key considerations involved, we use a stylized setting in this section with two types of managers:[13]

Good: producing returns with an (expected) annualized Sharpe ratio of 0.5; and
Bad: producing returns with an (expected) zero Sharpe ratio.

The central question we seek to answer is: what performance statistics are the most informative for deciding whether to replace a manager? We will first discuss a setting with a single decision moment, which reduces the analysis to a question of how well we can disentangle Good and Bad managers, based on different statistics. Next we explore a richer setting, with a monthly decision to replace a manager or not. In this case, it also matters how quickly one is able to detect (and replace) Bad managers. In the final part of this section, we look at time-varying drawdown threshold, which are more complex (perhaps explaining why they are not commonplace), yet more appropriate for the case.

Figure 5: Illustrative manager performance
Illustrative example of a manager's cumulative return with the maximum drawdown and total return since inception highlighted.



3.1 Classification at the end of a 10-year observation period

We need to recognize that the decision to replace a manager will be subject to two types of errors:

Type I error: replacing a Good manager,
Type II error: not replacing a Bad manager.

In Figure 6, we show the tradeoff between these two error types for three rules applied after a 10-year observation window:

(1) Total return over the 10 years;
(2) Drawdown level at the 10-year point;
(3) Maximum drawdown during the 10-year period.

take any more pain, see e.g. Thaler and Johnson (1990). This could make a drawdown all the more salient: after experiencing initial painful losses, people are then subjected to more, which is likely to be extra painful.

[13] To simplify the analysis, we abstract from adding a third "Ugly" type, sometimes considered in studies with heterogeneous agents.

Each dot in Figure 6 corresponds to a different cutoff value for the respective statistic. A larger diamond highlights the case where we use -10% (-1 annual standard deviation) as the cutoff value for each statistic.

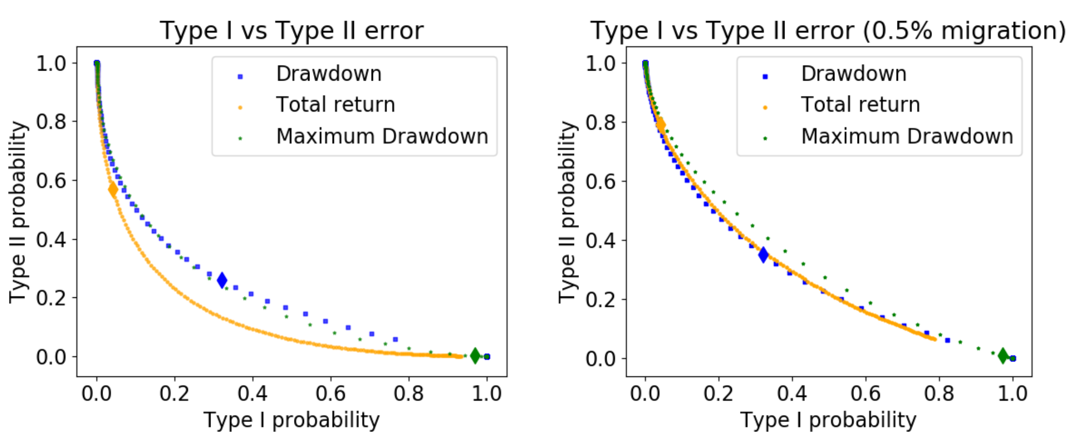
In the left panel of Figure 6, we assume we have a pool consisting of 50% Good (Sharpe ratio 0.5) and 50% Bad (Sharpe ratio 0.0) managers with returns that are normal, IID, and with an annualized standard deviation of 10%. So, the annualized mean return is 5% or 0%, depending on whether the manager is Good or Bad. Crucially, we assume managers are of constant type. Here, it is clear that classification based on the total return leads to a better Type I/Type II tradeoff than using a drawdown-based rule, as the curve is closer to the origin (low Type I and Type II errors). This should come as no surprise, as the only unknown of the manager return distribution is the mean return. The realized mean (or total) return is a sufficient statistic, using all historical returns with equal weight. In contrast, the statistics based on peak and/or trough returns are a complicated, path dependent function of historical returns.

In the right panel of Figure 6, we assume all managers start off as Good, but that they migrate to Bad at a constant monthly rate over time. The assumed monthly migration rate is 0.5%, which means that after 10 years, around 45% of managers have migrated from Good to Bad. These assumptions are motivated by the fact that in practice, managers or strategies can migrate from Good to Bad because of structural market changes, increased competition for the strategy style employed, staff turnover, or a fund accumulating too many assets. Now, the drawdown and total return based rules are similarly effective. This is a big change from the case of constant manager types (left panel), where the total return-based rule was superior. The pick-up in the appeal of drawdown-based rules here is intuitive, as they put more emphasis on recent history, and so are more tailored to the possibility of a migration from Good to Bad.

Figure 6: Efficacy classification rules with a 10-year horizon

We show the Type I error (mistakenly replacing a Good manager) and Type II error (mistakenly keeping a Bad manager) for three replacement rules. Evaluation takes place after observing 10 years of monthly data. In the left panel, the pool of managers consists of 50% Good and 50% Bad managers. In the right panel, all managers start off as Good, but each month there is a 0.5% chance of migrating to a Bad manager type. Good and Bad managers have a Sharpe ratio of 0.5 and 0.0 respectively. Returns are normal, IID, with 10% annualized volatility for both manager types. Different observations correspond to different cutoff values for the replacement rule, with a diamond corresponding to a -1 sigma cutoff.

No migration | **Migration probability of 0.5% per month**



As can be seen in the Figure 6, a -10% cutoff value (represented by the big diamonds in the plot) leads to very different Type I error values. It is, for example, much more common for the drawdown level to hit -10% than it is for the total return to reach -10%. In fact, every time the total return hits -10%, the drawdown must also be at least as bad as -10%. The reverse does not hold.

In order to do a sensible comparison, in Table 1, we report the Type I and II error rates across the three rules for a given implied probability of replacement (reported in the first column). The rules require different cutoff values in order to have the same probability of hitting the cutoff after the 10-year window (with no type migration). Consistent with the left panel of Figure 6, we see that the total return-based rule is preferred. The total-return rule is superior in terms of its lower Type 1 error (fewer Good managers are incorrectly identified as Bad).

Table 1: Cutoff values associated with a given probability of replacement after 10 years

For different probabilities of replacement, we tabulate the associated cutoff value for the three replacement rules considered, as well as the Type I and Type II error rate.

Probability of replacement (%)	Total Return			Drawdown			Maximum Drawdown		
	Threshold (%)	Type I	Type II	Threshold (%)	Type I	Type II	Threshold (%)	Type I	Type II
10	-27	0.01	0.80	-35	0.01	0.81	-41	0.01	0.82
20	-15	0.03	0.64	-26	0.05	0.63	-34	0.05	0.64
30	-3	0.07	0.47	-20	0.10	0.49	-30	0.10	0.51
40	9	0.13	0.33	-15	0.18	0.37	-26	0.19	0.37
50	22	0.22	0.21	-11	0.29	0.28	-23	0.29	0.26

3.2 Monthly evaluation

In reality, the decision to replace a manager is not done once, at the end of a long observation window, but also intermittently. For example, some multi-manager hedge funds state very clearly at what drawdown level a portfolio manager is fired. Interestingly, typically a constant cutoff value is used, rather than allowing for larger drawdowns when a manager has been running for a longer time. The reasons for this may be behavioral – in other words the rule is intended to alter manager behavior whether the manager has long tenure or not. It is also possible that the fund is acknowledging the difficulty of determining which managers are good at any point and allowing the drawdown rule to do the work for them recognizing that some good managers become bad. In this sub-section, we follow this practice and assume constant cutoff values. In the next sub-section, we will contrast a constant with a time-varying drawdown rule.

Concretely, we now consider an investor who evaluates managers monthly for a 10-year period. In Figure 7, we compare the efficacy of a total return and maximum drawdown rule to replace managers, where we make the same assumptions on manager types as before in Figure 6.[14] That is, in the left panel, managers are of constant type (50% Good, 50% Bad), while in the right panel, all managers start off as Good, but they migrate to a Bad type at a constant rate. Replacement

[14] Notice that in the case of monthly evaluation a maximum drawdown and a drawdown rule with the same cutoff value are equivalent This holds because the maximum drawdown is determined using monthly data, and the evaluation of the rule occurs at the same monthly frequency, so a maximum drawdown and drawdown rule will both cross a cutoff value at the same moment for the first time, and so trigger at the same time.

means drawing a new manager from the same pool, i.e., 50-50 odds of Good-Bad in the case of the left panel, and a Good manager (that can deteriorate subsequently) in the case of the right panel.

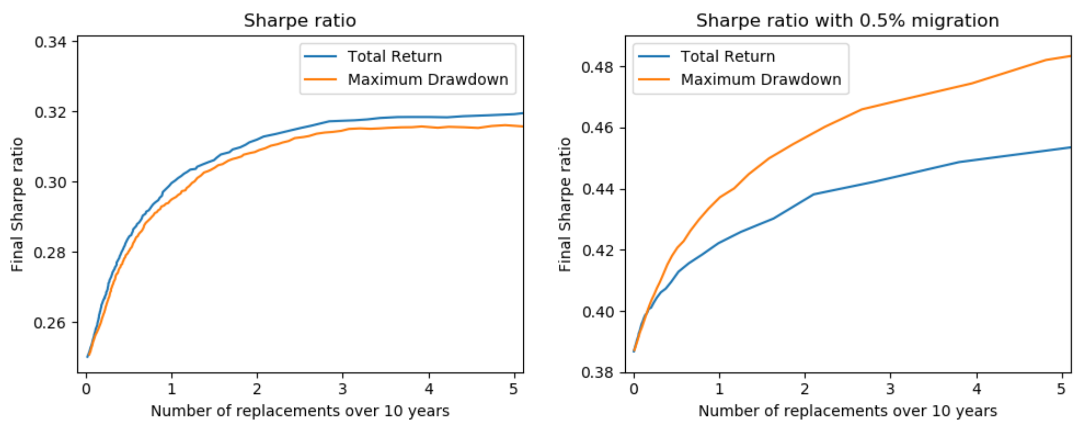
Just comparing Type I and II error rates is not sufficient in the case of monthly evaluation anymore, because it also matters how fast a bad manager is replaced. So instead, Figure 7 shows the Sharpe ratio over a 10-year window, with managers being replaced when they hit the threshold value.

Replacing a manager can be costly, e.g., because it requires due diligence into new managers, involves legal costs, and resets the high water mark in case of performance-fee charges for hedge funds. For this reason, in Figure 7, we plot the resulting Sharpe ratio when using the two replacement rules as a function of the average number of replacements during the 10-year window. To this end, we vary the cutoff value, and, for each value, plot the average Sharpe ratio as a function of the total number of replacements over the 10-year period.

In the left panel of Figure 7, we see that in case of constant manager types, the total return is better than the drawdown-based rule. This is consistent with the left panel of Figure 6. Again, the intuition is that the total return is an efficient statistic for estimating a manager's average return, while the drawdown statistic is path-dependent and so more wasteful in its use of historical return observations.

Figure 7: Efficacy of replacement rules with monthly evaluation

We show the average Sharpe ratio over a 10-year window, with a monthly decision to replace managers based on either a total return- or drawdown-based rule. In the left panel, the pool of managers consists of 50% Good and 50% Bad managers. In the right panel, all managers start off as Good, but each month there is a 0.5% chance of migrating to a Bad manager type. Good and Bad managers have a Sharpe ratio of 0.5 and 0.0 respectively. The average Sharpe ratios in the right panel are higher because of a greater proportion of Good managers. Monthly returns are normal, IID, with 10% annualized volatility for both manager types. We vary the cutoff value used in the replacement rule and plot the average Sharpe ratio against the average number of replacements.



In the right panel of Figure 7, we see that in case of a manager migrating from Good to Bad, a drawdown-based replacement rule is more effective in that it results in a higher Sharpe ratio for a given number of replacement over 10 years. The superior performance of the drawdown-based rule is intuitive, as it more naturally picks up on recent, sudden drop-offs in performance.

3.3 Monthly evaluation with a changing drawdown threshold

In practice, investors often employ fixed drawdown thresholds, even though the probability of hitting said value increases through time, as previously illustrated in Figure 2 (Panel B). In order to have a more consistent rate of replacement through time, we now also consider a drawdown cutoff value that increases with time. Specifically, the cutoff is:

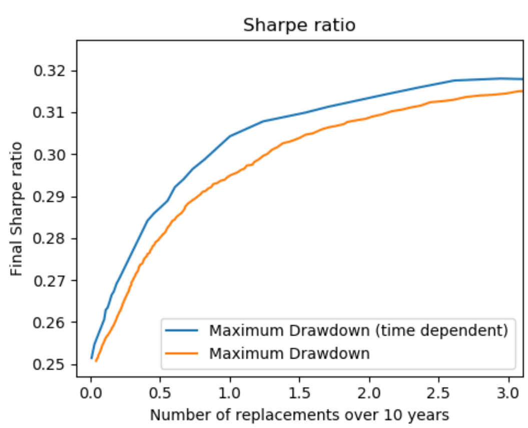
$$cutoff(t) = k \times max(1, \sqrt{t/12}),$$

for different threshold values, k, and number of months, t. The square-root term is motivated by the fact that the volatility of cumulative returns tends to grow approximately with the square root of time. We take the maximum of 1 and time/12, so that in the first year, we are not working with a very low threshold.

In Figure 8, we show the Sharpe ratio over a 10-year window for the time-varying drawdown rule versus the constant rule considered before. As before, in Figures 6 and 7 (left panels), replacement means drawing a new manager from a constant pool of 50% Good and Bad managers. The time-dependent rule leads to better performance, for a given number of replacements over the 10-year window. The effect seems to flatten off though with more frequent replacements, as the time effect is less relevant then.

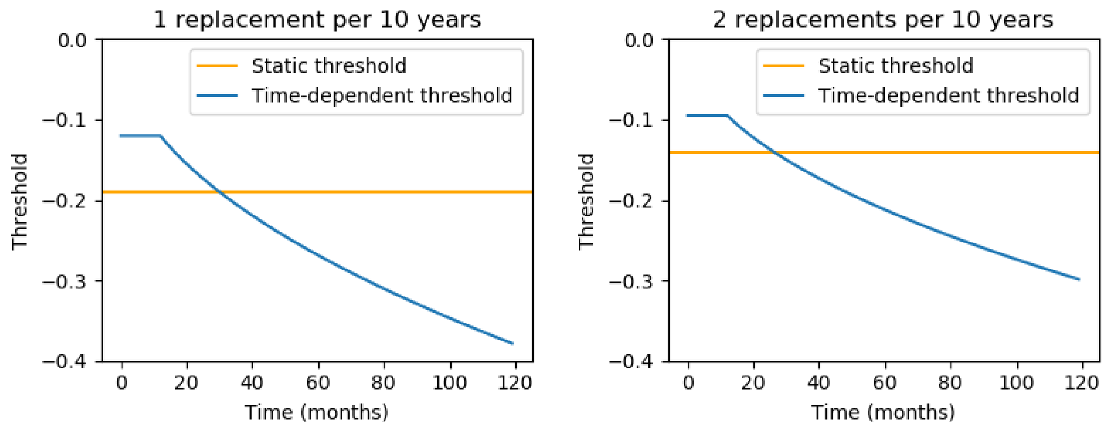
Figure 8: Efficacy of the drawdown replacement rules with monthly evaluation and a static or time-dependent threshold

We show the average Sharpe ratio over a 10-year window, with a monthly decision to replace managers based on a drawdown rule with either a stationary or time-dependent threshold. The pool of managers consists of 50% Good (Sharpe ratio 0.5) and 50% Bad (Sharpe ratio 0.0) managers. Monthly returns are normal, IID, with 10% annualized volatility for both manager types. We vary the cutoff value used in the replacement rule and plot the average Sharpe ratio against the average number of replacements.



In Figure 9, we show the static and time-varying drawdown thresholds in case of 1 and 2 replacements per 10 years on average in the left and right panel, respectively. The time-varying rule starts off with a significantly lower drawdown threshold, but is less stringent at longer horizons. This lines up better with the probability of drawdowns of a given level increasing with time (see Figure 2, Panel B).

Figure 9: Equivalent thresholds through time for the different drawdown replacement rules
We show the equivalent threshold through time used for each drawdown rule for one and two replacements over two years.



4. Drawdown-based risk reduction rules

In the previous section, we considered rules for replacing managers. Another common application of drawdown rules is to use them to first lower the risk of a manager while continuing to evaluate subsequent performance.

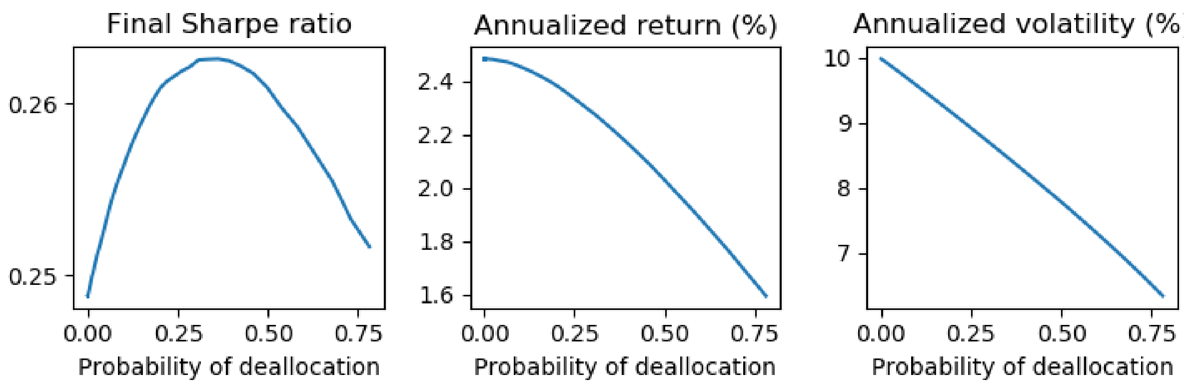
In Figure 10, we illustrate the effect of a drawdown-based rule, where a risk reduction of 50% is triggered if the drawdown dips below a cutoff value. Full risk taking is restored if the manager recoups half of the losses. That is, they would have recovered the peak-to-trough loss if their risk had not been reduced by half. We vary the cutoff used and plot the Sharpe ratio, annualized return, and annualized volatility against the probability of having at least one risk reduction over the 10-year evaluation window. While the Sharpe ratio can improve slightly from such a risk reduction rule, the annualized return is lower. This is perhaps an obvious result, as there is always still a chance the manager is Good (0.5 Sharpe ratio). In the worst case that they are Bad, it has a zero (and not negative) Sharpe ratio.

While this illustration may be obvious, it shows that risk reductions may only serve to reduce risk, and not improve the annualized return, unless one takes up risk elsewhere in the portfolio. Of course, if the pool of managers is finite or there are costs to taking on new managers, then a rule like this might serve a practical purpose.

For a risk reduction method to improve average returns, the conditional expected returns need to be negative. This can happen if there is a large chance of having a negative Sharpe ratio manager, but this seems a stretch, as it requires negative skill or very high transaction costs. An alternative is a setting with a very high degree of autocorrelation, where one may have a negative expected return following a negative realized return.[15]

[15] It can be shown that a stop loss policy adds value when the level of serial autocorrelation in an AR(1) process is greater than the Sharpe of the process, see Kaminski and Lo (2014).

Figure 10: Efficacy of risk reduction rule, monthly evaluation, no type migration
We show the average Sharpe ratio over a 10-year window, with a monthly decision to reduce risk based on a drawdown rule. We simulate a pool of managers with a 50% change of being Good (Sharpe ratio 0.5) or Bad (Sharpe 0.0).



5. Five key takeaways

So, what type and size of drawdown should cause you to change an investment manager? We offer five main conclusions, presented here in the order in which they are discussed in the paper.

First, know your stats. Drawdowns are easy to compute. However, it is challenging to estimate the probability of hitting a certain drawdown level. As such, we help you set sensible drawdown limits for given (or stated) parameters of the return distribution.

Second, a pre-set drawdown rule may prevent peak risk taking. Taking risk in bursts (leading to heteroscedastic or kurtotic returns) will increase the probability of hitting a certain drawdown level, relative to more constant risk taking (holding constant the long-term volatility). Hence, clearly communicated drawdown limits can motivate a manager to take more even risk over time. Also, automatic de-allocation at a given drawdown level may prevent a manager from adverse behavior to exploit the 'trader put', i.e., increasing risk when returns fall to maximize chances of recovery, while further losses do not cause the manager much further pain due to the limited liability nature. Similarly, a deep out-of-the money put is only very valuable with a high degree of volatility.

Third, think in terms of the relative cost of Type I and Type II errors; see also Harvey and Liu (2020). If hiring a manager is a costly endeavor, Type I errors (booting good managers) are costly. If a bad manager just adds noise (has a Sharpe ratio of zero) in an otherwise diversified portfolio, and if ample cash is available, some Type II errors (keeping a bad manager) may not be that bad. However, if bad managers have a negative Sharpe ratio (e.g., because of transaction costs, or because they unwittingly take the other side of the trade of some shrewd investors), Type II errors become much more of a concern. Thinking in terms of the costs of Type I and Type II errors is crucial for the hiring and firing process.

Fourth, look at both total return and drawdown statistics. Total-return (or Sharpe-ratio) rules are best at measuring the constant ability of a manager to create positive returns. Drawdown-based rules, on the other hand, are better suited to deal with a situation where a manager abruptly loses their skill. In reality, the two are complementary, where the relative weight on total returns versus drawdown depend crucially on the assessment of how likely it is that a good manager can transition into a bad one. Obviously, alternative criteria that may hint at a possible deterioration of a manager's quality (such as turnover, fast asset growth, publication of their secret sauce) may provide a warning that an investor should start to place more weight on the drawdown statistic.

Fifth, consider a time-varying drawdown rule. The probability of hitting a certain drawdown level naturally increases over time, even if a manager continues to be of the same type, generating returns from a constant distribution. Somewhat puzzlingly, drawdown limits in practice are typically set at a constant (time-invariant) level. Even though it adds some complexity, a time-varying drawdown rule is advisable.

References

Bailey, D. and M. López de Prado (2015), "Stop-outs under serial correlation and the triple penance rule", *Journal of Risk*, 18(2): 61-93.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012), "Salience theory of choice under risk", *Quarterly Journal of Economics*, 127(3): 1243–1285.

Busseti, E., E. Ryu, and S. Boyd (2016), "Risk-constrained Kelly gambling", *Journal of Investing*, 25(3): 118-134.

Carr, P., H. Zhang and O. Hadjiliadis (2011), "Maximum Drawdown Insurance", *International Journal in Theoretical and Applied Finance*, 8(14): 1195–1230.

Casati, A. and S. Tabachnik (2012), "The statistical properties of the maximum drawdown in financial time series", working paper.

Chekhlov, A., S. P. Uryasev and M. Zabarankin (2005), "Drawdown Measure in Portfolio Optimization", I*nternational Journal of Theoretical and Applied Finance*, 8(1): 13-58.

Cvitanic, J., S. Kou, X. Wan, and K. Williams (2019), "Pi portfolio management: reaching goals while avoiding drawdowns", working paper.

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

Efron, B. and R. Tibshirani (1986), "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy", Statistical Science, 1(1): 54-75.

Goyal, A. and S. Wahal (2008), "The selection and termination of investment management firms by plan sponsors", *Journal of Finance*, 63(4), 1805-1847.

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

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

Harvey, C. R., E. Hoyle, R. Korgaonkar, S. Rattray, M. Sargaison and O. Van Hemert (2018) "The impact of volatility targeting", *Journal of Portfolio Management*, 45(1): 14-33.

Harvey, C. R., and Y. Liu, (2020), "False (and missed) discoveries in financial economics, *Journal of Finance*, forthcoming.

Kaminski, K. M. and A. W. Lo (2014), "When do Stop-Loss Rules Stop Losses?" *Journal of Financial Markets*, 18: 234–254.

Korn, O., P. M. Möller and C. Schwehm (2020), "Drawdown Measures: Are They All the Same?" working paper.

Leal, R. P. C. and B. V. de M. Mendes (2015), "Maximum Drawdown: Models and Applications", *Journal of Alternative Investments,* 7(4): 83-91.

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

Molyboga, M., C. L'Ahelec (2017), "Portfolio Management with Drawdown-Based Measures", *Journal of Alternative Investments*, 19(3):75-89.

Sornette, D. (2003), *Why stock markets crash: critical events in complex financial systems*. Princeton University Press, Princeton.

Thaler, R., and E. Johnson (1990), "Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice", *Management Science*, 36(6): 643-660.

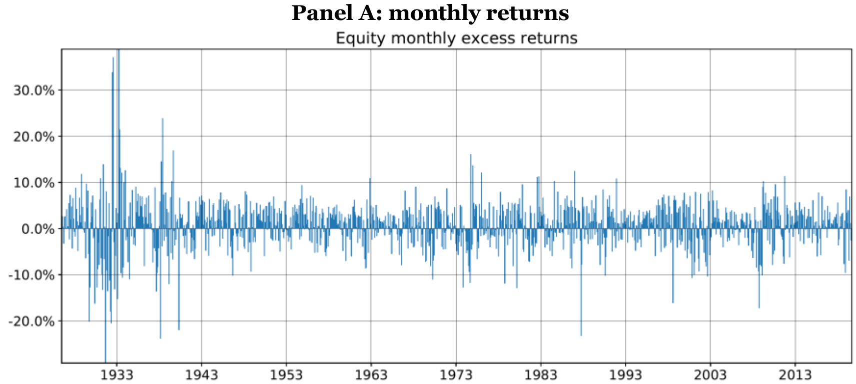
Vecer, J. (2006), "Maximum Drawdown and Directional Trading," *Risk*, 19(12), 88–92.

Appendix A. Heteroscedasticity for US stocks

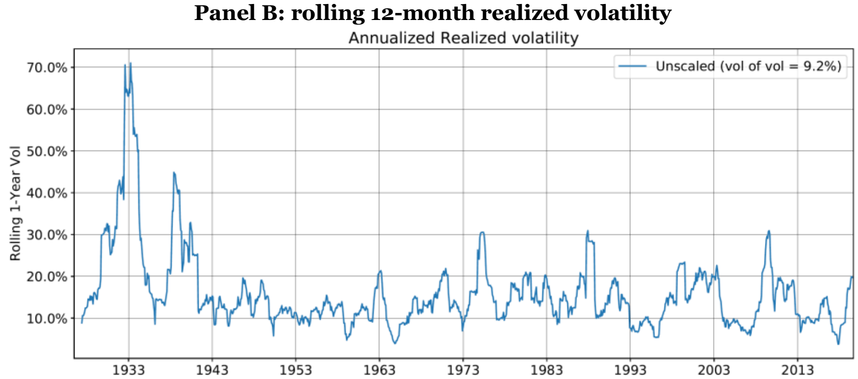
In Figure C1 we show monthly US equity returns (panel A) and the rolling 12-month realized volatility for US equity returns from June 1926 to December 2019. Volatility is persistently high around, for example, 1929 (Great Depression) and 2008 (Global Financial Crisis).[16]

.

Panel A: monthly returns



Panel B: rolling 12-month realized volatility



[16] From K. French's website:
https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Appendix B. Historical gap moves

Table B1: Historical biggest negative (gap) moves

The table shows for a number of liquid securities the sample period (start date, end date, total number of years) and the worst monthly return (which month, percent return, number of annual standard deviations).

	Sample period		Most negative return				
Description	Start	End	Month	Value	#STDs (ann)	Vol of Vol (re-scaled to 10%)	Autocorrelation
EQUITY INDICES							
CAC 40 Index	Nov 1988	Feb 2020	Aug 1990	-16.6%	-1.27	5.6%	0.09
DAX Index	Nov 1990	Feb 2020	Sep 2002	-23.0%	-1.64	6.2%	0.05
NASDAQ 100 Index	Apr 1996	Feb 2020	Feb 2001	-28.8%	-1.85	7.2%	0.05
Russell 2000 Index	Sep 2000	Feb 2020	Oct 2008	-20.5%	-1.88	6.5%	0.05
S+P 500 Index	Apr 1982	Feb 2020	Oct 1987	-20.4%	-1.77	3.5%	0.04
Euro-STOXX	Jun 2000	Feb 2020	Sep 2002	-18.0%	-1.76	6.8%	0.10
FTSE	May 1984	Feb 2020	Oct 1987	-27.6%	-2.30	4.3%	-0.01
Hang Seng	Jan 1987	Feb 2020	Oct 1987	-40.7%	-2.21	5.2%	0.02
Korean Kospi	Sep 2000	Feb 2020	Sep 2001	-23.3%	-1.85	7.3%	0.02
Nikkei	Mar 1987	Feb 2020	Oct 1987	-32.8%	-2.15	5.1%	-0.01
Average				-25.2%	-1.87	5.8%	0.04
GOVERNMENT BONDS							
German Bonds	Mar 1997	Feb 2020	Jan 2011	-0.9%	-1.31	7.0%	0.20
German Bonds	Jun 1983	Feb 2020	Feb 1990	-6.1%	-1.53	4.1%	0.05
Gilts	Nov 1982	Feb 2020	Sep 1986	-9.4%	-1.60	3.7%	0.05
Japanese Bonds	Mar 1983	Feb 2020	Sep 1987	-7.2%	-1.84	4.7%	0.01
German Bonds	Oct 1991	Feb 2020	Feb 1994	-1.9%	-0.96	6.0%	0.14
US Treasuries	Sep 1977	Feb 2020	Jul 2003	-9.4%	-1.06	1.7%	0.05
US Treasuries	Jul 2005	Feb 2020	Apr 2008	-1.1%	-1.95	6.8%	0.22
US Treasuries	Oct 1991	Feb 2020	Apr 2004	-3.1%	-1.15	6.1%	0.12
US Treasuries	May 1982	Feb 2020	Jul 2003	-5.6%	-1.08	3.2%	0.05
Average				-5.0%	-1.39	4.8%	0.10
OIL							
Crude Oil	Jun 1988	Feb 2020	Oct 2008	-37.3%	-1.51	5.4%	0.26
Crude Oil	Oct 1983	Feb 2020	Oct 2008	-35.9%	-1.47	4.5%	0.17
Heating Oil	Mar 1979	Feb 2020	Oct 2008	-32.7%	-1.29	2.7%	0.09
Gas Oil	Apr 1981	Feb 2020	Oct 2008	-30.7%	-1.21	4.0%	0.21
RBOB Gasoline	Dec 1984	Feb 2020	Oct 2008	-41.1%	-1.55	4.6%	0.07
Average				-35.5%	-1.41	4.2%	0.16
METALS							
Aluminium	Jan 1980	Feb 2020	Sep 1988	-20.0%	-1.17	3.2%	0.02
Copper	Jul 1959	Feb 2020	Oct 2008	-36.3%	-1.37	0.4%	0.11
Gold	Jan 1975	Feb 2020	Mar 1980	-25.4%	-1.60	1.6%	-0.05
Lead	Jun 1989	Feb 2020	May 2008	-29.9%	-1.59	5.7%	0.04
Nickel	Jul 1979	Feb 2020	Oct 2008	-27.3%	-0.98	3.6%	0.13
Silver	Jun 1963	Feb 2020	Mar 1980	-45.6%	-1.62	1.3%	0.01
Zinc	Jan 1975	Feb 2020	Oct 2008	-34.0%	-1.67	2.0%	0.01
Average				-31.2%	-1.43	2.5%	0.04