

Drawdowns

Otto Van Hemert, Mark Ganz, Campbell R. Harvey,
Sandy Rattray, Eva Sanchez Martin,
and Darrel Yawitch

OTTO VAN HEMERT
is the head of macro research
at Man AHL in London,
UK.
otto.vanhemert@man.com

MARK GANZ
is a quantitative researcher
working in London, UK.
ganzmark@gmail.com

CAMPBELL R. HARVEY
is professor of finance
at Duke University in
Durham, NC, and an
advisor to Man Group in
London, UK.
cam.harvey@duke.edu

SANDY RATTRAY
is the CIO of Man Group in
London, UK.
sandy.rattray@man.com

EVA SANCHEZ MARTIN
is a quantitative researcher at
Man AHL in London, UK.
eva.sanchezmartin@man.com

DARREL YAWITCH
is the chief risk officer at
Man AHL, Man GLG, and
Man Solutions in London,
UK.
darrel.yawitch@man.com

*All articles are now
categorized by topics
and subtopics. **View at**
PM-Research.com.

KEY FINDINGS

- Drawdowns are particularly useful for detecting managers who have lost their ability to outperform. It is important to look at both drawdowns and overall performance measures such as Sharpe ratios.
- Drawdown limits need to be dynamically adjusted to control for the length of time the assets have been managed. There is a greater chance for bad luck when the duration is longer.
- When interpreting realized drawdowns, it is crucial to take the relative costs of Type I errors (firing good managers) and Type II errors (keeping bad managers) into account. Type I errors are particularly expensive if it is costly to find a new manager.

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, the authors begin by quantifying the probability of hitting a certain drawdown level, given various return distribution properties. Next, they 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, they show that drawdown-based rules can be used as a risk reduction technique, but this affects both expected returns and risk.*

TOPICS: *Portfolio theory, portfolio construction, manager selection, wealth management**

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 because its calculation is unambiguous and it 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 such as skewness or semivariance 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.¹ In the first part of this article, 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 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 nonnormal, but still time-independent, returns (e.g., 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: Multiperiod 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.² We also recognize that the timing of these replacements matters because bad managers 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 many assets (decreasing returns to scale).

¹ See, for example, Magdon-Ismail et al. (2004), who studied the behavior of maximum drawdown for the case of a Brownian motion with drift and derived an analytic expression for the expected value of maximum drawdown (with zero drift) and infinite series representation (for nonzero drift).

² See also Harvey and Liu (2020) for an analysis of the trade-off between Type I and Type II errors, as well as their differential costs.

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—unless a manager's conditional expected 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 to believe there is a very high degree of persistence in returns and that previous returns were negative.

Finally, we summarize the main results and discuss five key takeaways for allocators choosing among managers or for managers choosing among 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 a “poor man's volatility scaling.”³

DRAWDOWN GREEKS

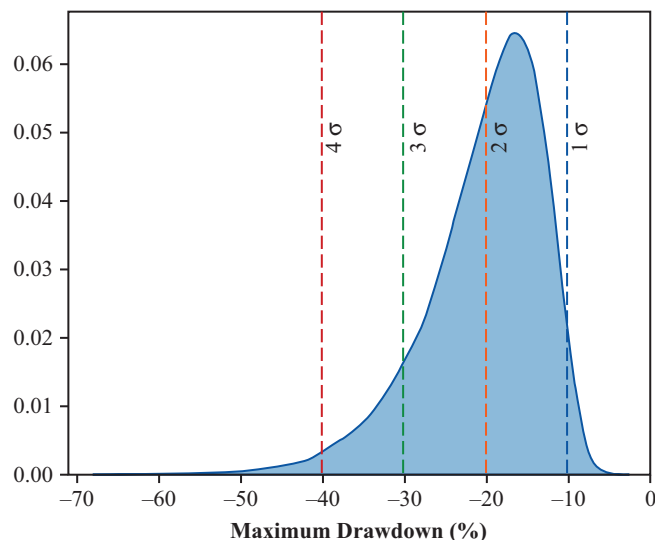
In this section, we explore how sensitive the likelihood of hitting a certain drawdown level is to key drivers, such as the Sharpe ratio, evaluation time window, and autocorrelation of returns. Borrowing terminology from option pricing theory, we call these sensitivities the *drawdown Greeks*.⁴

³ See Harvey et al. (2018) for a discussion of the impact of volatility scaling on risk and return characteristics.

⁴ 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 López de Prado 2015; and Busseti, Ryu, and Boyd 2016); (2) portfolio optimization (see, e.g., Grossman and Zhou 1993; Chekhlov, Uryasev, and Zabarankin 2005; and Cvitanic et al. 2019); (3) hedging and risk management (see, e.g., Carr, Zhang, and Hadjiliadis 2011; Leal and Mendes 2015; and Molyboga and L'Ahelec 2017); (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).

Exhibit 1

Probability Distribution for the Maximum Drawdown Statistic



Notes: The probability distribution for the maximum drawdown statistic using normal, i.i.d. 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 one, two, three, and four annual standard deviations.

Source: Man AHL.

Probability Distribution for Maximum Drawdown Level

We start with a simple setting of normal independent and identically distributed (i.i.d.) monthly returns.⁵ In Exhibit 1, we show the probability distribution of the maximum drawdown statistic for our baseline case: 10-year time window, 10% annualized volatility, and 0.5 annualized Sharpe ratio. Each parameterization is evaluated with 100,000 simulations of monthly returns for the evaluation window.

⁵ 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 because they are influenced by a pronounced intramonth variation in the news flow (e.g., bigger moves on the days major economic news is released). Monthly returns are somewhat better behaved because 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.

We highlight with vertical lines maximum drawdown levels of one, two, three, and four 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 or 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 one, two, three, and four sigma levels, respectively. Thus, 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. In 1 in 10 cases, one even reaches a drawdown of three full annual standard deviations (or -30%).

Drawdown Greeks without Higher Order Effects

Next, we consider how deviations from the baseline assumptions affect the probability of hitting a drawdown level. In Exhibit 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: (Panel A) annualized volatility, 10% baseline; (Panel B) time window, 10-year baseline; (Panel C) annualized Sharpe ratio, 0.5 baseline; and (Panel D) autocorrelation, 0.0 baseline.⁶

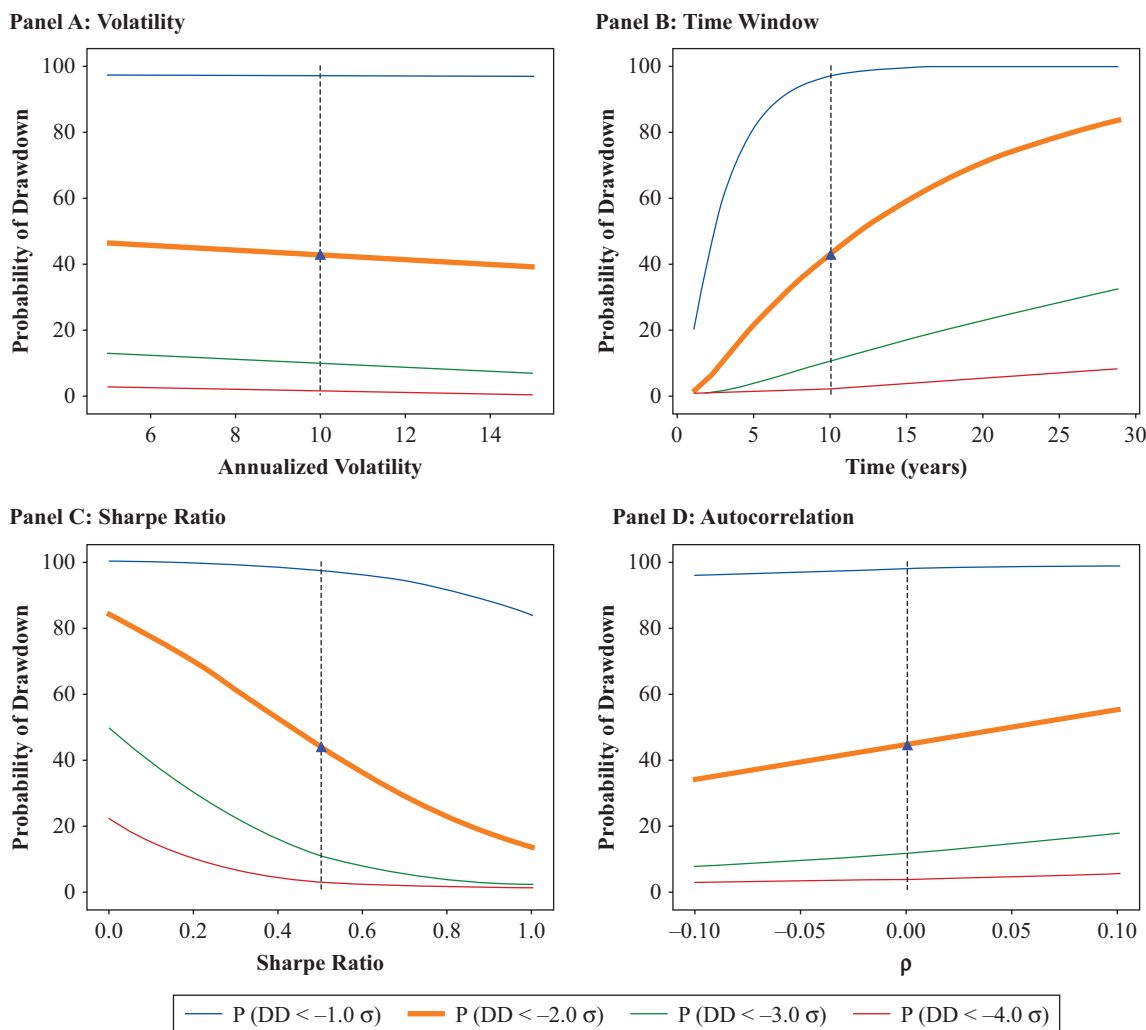
In Panel A, we show how the probability of a given maximum drawdown changes when we vary the standard deviation of the return process while holding constant the Sharpe ratio. The orange line represents the probability of a maximum drawdown that is -2 sigma (annual return standard deviations) or worse.⁷ This value is 43% for the baseline case (see also the discussion of Exhibit 1), indicated by the vertical dashed line. The orange line is near horizontal; that is, varying the standard deviation of returns hardly changes the probability of reaching a certain maximum drawdown level, as long as one assumes the Sharpe ratio stays constant and the threshold is expressed in terms

⁶ Bailey and López 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.

⁷ 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.

Exhibit 2

Sensitivity of the Probability of a Maximum Drawdown to Key Parameters



Notes: The probability of reaching a maximum drawdown of one, two, three, and four sigma (annual standard deviations). In all cases, returns are normal and identically distributed. In the different panels, we vary the (Panel A) volatility while holding constant the Sharpe ratio, (Panel B) time window, (Panel C) Sharpe ratio, and (Panel 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.

Source: Man AHL.

of sigmas. That is, the probability of a 20% maximum drawdown when returns have a 10% standard deviation is similar to the probability of a 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 also increase as volatility increases). The orange line is nearly horizontal, but not exactly. In fact, it gently slopes downward, reflecting the influence of the compounding of returns.

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 constant the standard deviation of returns. In the default case, we have an annualized Sharpe ratio of 0.5. The impact of the Sharpe ratio on the probability of reaching

a certain maximum drawdown level is large, which is intuitive because 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 following formula, the $\tilde{\mu}$ and $\tilde{\sigma}$ terms capture the unconditional mean and standard deviation, respectively; we use a tilde to clarify that it concerns monthly returns (in contrast to, e.g., Exhibit 2, in which we used σ , without a tilde, for the annualized standard deviation). The mean and standard deviations are premultiplied with a term featuring ρ to offset the effect of nonzero 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 i.i.d.

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 reasonable range because it implies a measurable degree of predictability.⁸ 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 .

Bootstrapped US Equity Returns

Next, we bootstrap two-year blocks from US equity returns since 1926 with monthly returns scaled to have 10% unconditional volatility.⁹ Using actual

⁸To motivate this statement, consider the monthly process defined earlier and set $\tilde{\mu} = 0$ for simplicity. If $R_t = +\tilde{\sigma}$, then 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 $\rho = 0.1$. If $R_t = +2\tilde{\sigma}$, the annualized conditional Sharpe ratio is 0.70 . These are higher than a typical annualized Sharpe ratio for general equities (say $8\%/20\% = 0.4$) or a typical hedge fund (with a Sharpe ratio of, say, 0.5 – 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 measurable degree of predictability.

⁹This is an approach in which one randomly selects (with replacement) blocks of consecutive observations from an actual distribution. As a robustness check, we reran our analysis with blocks

return realizations allows us to determine whether the inference is different from our simulated, normally distributed returns. Selecting blocks, rather than individual months, is done to preserve the original time-series structure within a block

In Panel A of Exhibit 3, we present the sensitivity to the time window, holding the Sharpe ratio constant at 0.5 (by adjusting the mean returns appropriately). In Panel B of Exhibit 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 Exhibit 2, in which 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 two-sigma drawdown (see Exhibit 2, baseline case for our simulated returns). This probability increases to 55% for the bootstrapped actual returns in Exhibit 3 (see dashed line crossing orange line). This increased probability of hitting a drawdown level is a result of both nonnormality of monthly returns and heteroskedasticity (clustering of volatility). This is illustrated in Appendix A for US equity returns since 1926, with volatility being persistently high around, for example, 1929 (the Great Depression) and 2008 (the Global Financial Crisis).

Although nonnormality and volatility clustering tend to increase the probability of hitting a certain drawdown level, the sensitivity to the time window and Sharpe ratio looks very similar between the normal (Exhibit 2) and nonnormal, bootstrapped case (Exhibit 3).

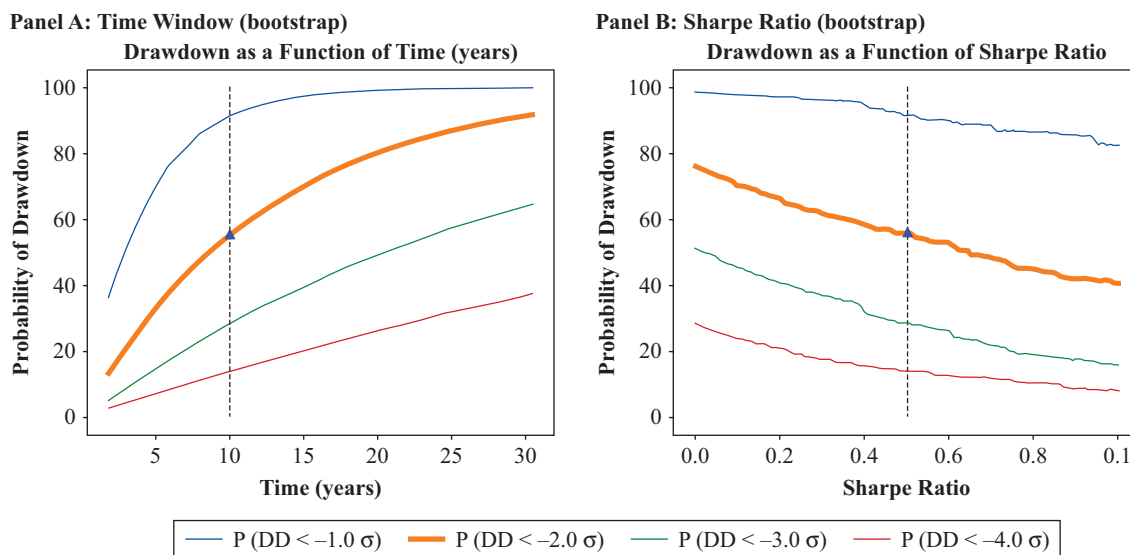
The Impact of Gap Risk

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

longer than 24 months and found similar results. See Efron and Tibshirani (1986) for an early discussion of bootstrap methods.

Exhibit 3

Drawdown Probability with Bootstrapped US Equity Returns



Notes: The probability of reaching a maximum drawdown of one, two, three, and four sigma (annual standard deviations), as a function of the time window (Panel A) and Sharpe ratio (Panel B). We bootstrap two-year blocks from US equity returns since 1926. We hold constant the unconditional volatility at 10%. In Panel A, we hold constant the Sharpe ratio at 0.5. In Panel B, we hold constant the time window at 10 years.

Source: Man AHL and Ken French's data library (mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

standard deviations (which corresponds to 3.5 to 5.2 monthly standard deviations).¹⁰

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 i.i.d.:

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

In Exhibit 4, we vary the size of the gap move (k in the preceding formula). The baseline corresponds to $k = 0$. Note that $k = 1$ is already a large value. In Appendix B, we show 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. Although the probability of a large drawdown indeed increases with an increased probability of a gap move, the impact is somewhat limited for $k = 1$ (monthly move equal to one annual standard deviation, or 3.5 monthly standard deviations). We think this is intuitive because such a move does not immediately take you through a, say, -2 sigma drawdown limit. Additionally, the drivers of returns in the medium to long term, such as the Sharpe ratio, remain the key drivers of the probability of hitting a drawdown.

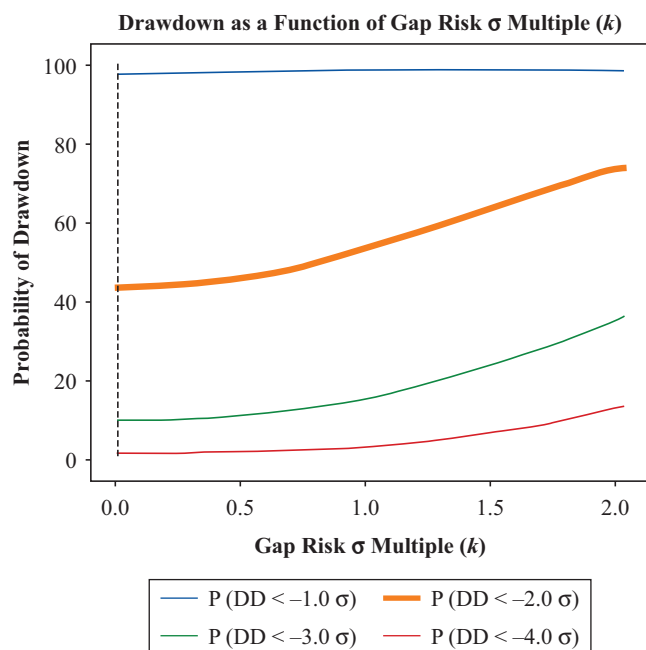
MANAGER REPLACEMENT RULES

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 of how to deal with a situation like the one illustrated in Exhibit 5, in which the total return for a manager still looks quite healthy

¹⁰ Under a normal distribution, a -5 sigma or worse monthly move happens less than once every 250,000 years.

Exhibit 4

Impact of Gap Risk



Notes: The probability of reaching a maximum drawdown of one, two, three, and four sigma (annual standard deviations). In the baseline case (indicated by the vertical dashed line), we have normal, i.i.d. 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.

Source: Man AHL.

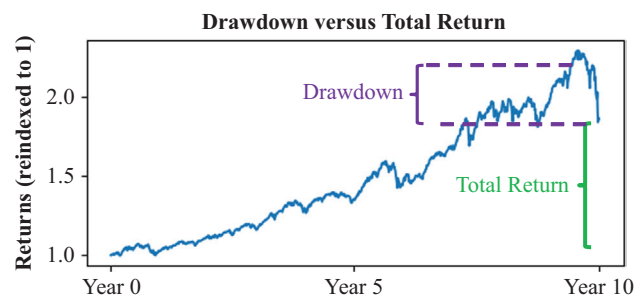
but the recent drawdown looks worrisome. Did something change? Was the manager never good in the first place?

To navigate the uncertainty around manager quality, investors need a framework for deciding whether to replace a manager. Otherwise, behavioral biases can lead to suboptimal decisions, as illustrated by Goyal and Wahal (2008), for example, who showed that investors are too quick to hire and fire managers.¹¹

¹¹ Behavioral biases may arise because drawdowns can be attention grabbing when observed in a graph like Exhibit 5. Such an effect has its foundation in the salience theory (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 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 loss, which is likely to be extra painful.

Exhibit 5

Illustrative Manager Performance



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

Source: Man AHL.

To illustrate the key considerations involved, we use a stylized setting in this section with two types of managers:¹²

- *Good*: producing returns with an (expected) annualized Sharpe ratio of 0.5
- *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 of whether to replace a manager. 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 thresholds, which are more complex (perhaps explaining why they are not commonplace) yet more appropriate for the case.

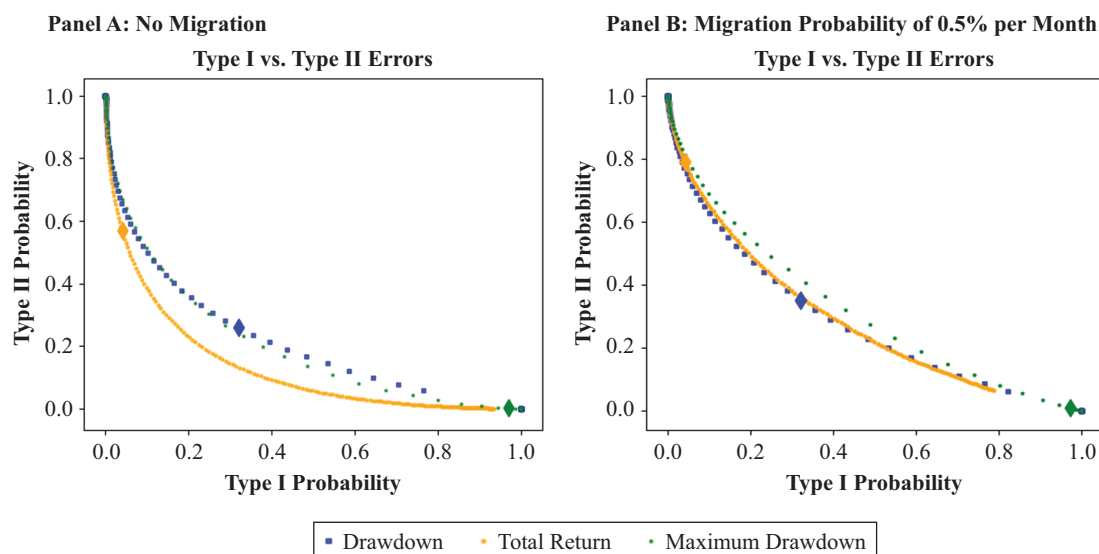
Classification at the End of a 10-Year Observation Period

It is important to recognize that the decision to replace a manager will be subject to two types of errors:

¹² To simplify the analysis, we abstract from adding a third *ugly* type, sometimes considered in studies with heterogeneous agents.

Exhibit 6

Efficacy Classification Rules with a 10-Year Horizon



Notes: The Type I errors (mistakenly replacing a good manager) and Type II errors (mistakenly keeping a bad manager) for three replacement rules. Evaluation takes place after observing 10 years of monthly data. In Panel A, the pool of managers consists of 50% good and 50% bad managers. In Panel B, all managers start off as good but each month have a 0.5% chance of migrating to bad. Good and bad managers have a Sharpe ratio of 0.5 and 0.0, respectively. Returns are normal, i.i.d., 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. A curve closer to the origin is better (smaller errors).

Source: Man AHL.

- **Type I error:** replacing a *good* manager
- **Type II error:** not replacing a *bad* manager

In Exhibit 6, we show the trade-off 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

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

In Panel A of Exhibit 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, i.i.d., and have an annualized standard deviation of 10%. Thus, the annualized mean return is 5% or 0%, depending on whether the manager is *good* or *bad*. Crucially, we assume managers are of a constant type.

Here, it is clear that classification based on the total return leads to a better Type I/Type II trade-off than using a drawdown-based rule because the curve is closer to the origin (low Type I and Type II errors). This should come as no surprise because the only unknown in 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 Panel B of Exhibit 6, we assume all managers start off as *good* but 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

Exhibit 7

Cutoff Values Associated with a Given Probability of Replacement after 10 Years

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

Notes: For different probabilities of replacement, the associated cutoff value for the three replacement rules considered, as well as the Type I and Type II error rates. No migration is assumed (analogous to Panel A of Exhibit 6).

Source: Man AHL.

a big change from the case of constant manager types (Panel A), where the total return–based rule was superior. The pickup in the appeal of drawdown–based rules here is intuitive because these rules put more emphasis on recent history and so are more tailored to the possibility of a migration from *good* to *bad*.

As can be seen in Exhibit 6, a −10% cutoff value or minus one standard deviation (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 poor as −10%. The reverse does not hold.

To improve the comparison, in Exhibit 7, 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 to have the same probability of hitting the cutoff after the 10-year window (with no type migration). Consistent with Panel A of Exhibit 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*).

Monthly Evaluation

In reality, the decision to replace a manager is not made once, at the end of a long observation window, but also intermittently. For example, some multimanager hedge funds state very clearly the drawdown level at which a portfolio manager will be fired. Interestingly, typically a constant cutoff value is used, rather than

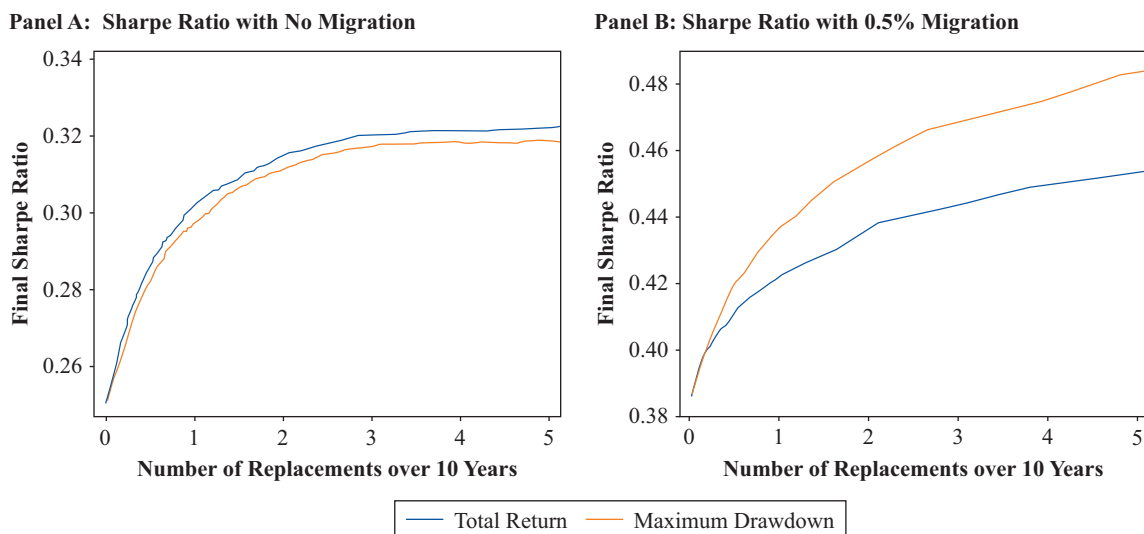
allowing for larger drawdowns when a manager has been running for a longer time. The reasons for this may be behavioral—in other words, the rule is intended to alter manager behavior whether the manager has long tenure or not. It is also possible that the fund is acknowledging the difficulty of determining 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 subsection, we follow this practice and assume constant cutoff values. In the next subsection, 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 Exhibit 8, we compare the efficacy of a total-return and maximum drawdown rule to replace managers, in which we make the same assumptions on manager types as before in Exhibit 6.¹³ In Panel A, managers are of constant type (50% *good*, 50% *bad*), whereas in Panel B, all managers start off as *good* but migrate to *bad* at a constant rate. Replacement means drawing a new manager from the same pool—that is, 50–50 odds of *good*–*bad* in the case of Panel A and a *good* manager (that can deteriorate subsequently) in the case of the Panel B.

¹³ 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. Thus, a maximum drawdown and drawdown rule will both cross a cutoff value for the first time at the same moment and so trigger at the same time.

Exhibit 8

Efficacy of Replacement Rules with Monthly Evaluation



Notes: 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 Panel A, the pool of managers consists of 50% good and 50% bad managers. In Panel B, all managers start off as good but each month have a 0.5% chance of migrating to bad. Good and bad managers have a Sharpe ratio of 0.5 and 0.0, respectively. The average Sharpe ratios in Panel B are higher because of a greater proportion of good managers. Monthly returns are normal, i.i.d., with 10% annualized volatility for both manager types. We vary the cutoff value used in the replacement rule (between 5% and 30%) and plot the average Sharpe ratio against the average number of replacements.

Source: Man AHL.

Just comparing Type I and II error rates is no longer sufficient in the case of monthly evaluation; it also matters how fast a *bad* manager is replaced. So instead, Exhibit 8 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 on new managers, involves legal costs, and resets the high-water mark in the case of performance-fee charges for hedge funds). For this reason, in Exhibit 8, 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 Panel A of Exhibit 8, we see that in the case of constant manager types, the total return is somewhat better than the drawdown-based rule. This is consistent with Panel A of Exhibit 6. Again, the intuition is that the total return is an efficient statistic for estimating a manager's average return, whereas the drawdown sta-

tistic is path dependent and so more wasteful in its use of historical return observations.

In Panel B of Exhibit 8, we see that in the event 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 replacements over 10 years. The superior performance of the drawdown-based rule is intuitive: It more naturally picks up on recent, sudden drop-offs in performance.

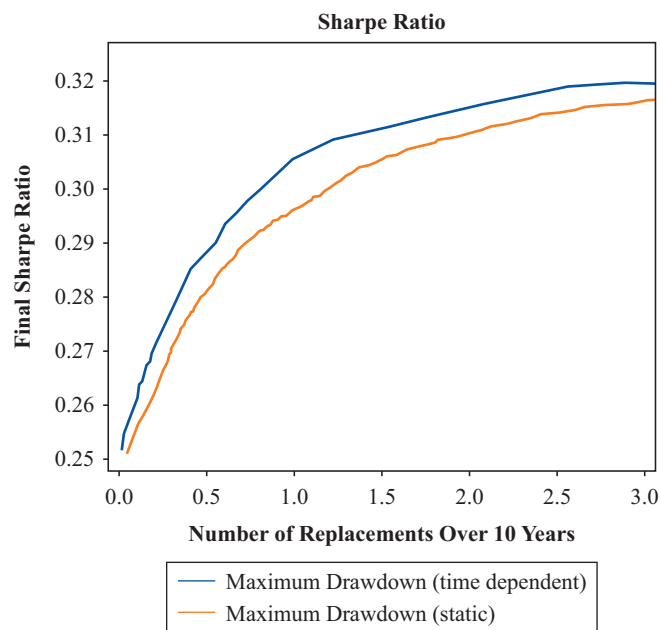
Monthly Evaluation with a Changing Drawdown Threshold

In practice, investors often employ fixed drawdown thresholds, even though the probability of hitting the said value increases through time, as previously illustrated in Exhibit 2 (Panel B). 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

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

Exhibit 9

Efficacy of the Drawdown Replacement Rules with Monthly Evaluation and a Static or Time-Dependent Threshold



Notes: 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, i.i.d., with 10% annualized volatility for both manager types. We vary the cutoff value used in the replacement rule (between 5% and 30%) and plot the average Sharpe ratio against the average number of replacements.

Source: Man AHL.

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 $\text{time}/12$ so that, in the first year, we are not working with a very low threshold.

In Exhibit 9, we show the Sharpe ratio over a 10-year window for the time-varying drawdown rule versus the constant rule considered earlier. As before, in Exhibits 6 and 8 (Panel A), 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 because the time effect is less relevant.

In Exhibit 10, we show the static and time-varying drawdown thresholds in the case of one and two replacements per 10 years, on average, in Panel A and Panel B, respectively. The time-varying rule starts with a much smaller drawdown threshold, but it is less stringent at longer horizons. This aligns better with the probability of drawdowns of a given level increasing with time (see Exhibit 2, Panel B).

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 reduce the allocation to the manager (reduce risk) the risk of a manager while continuing to evaluate subsequent performance.

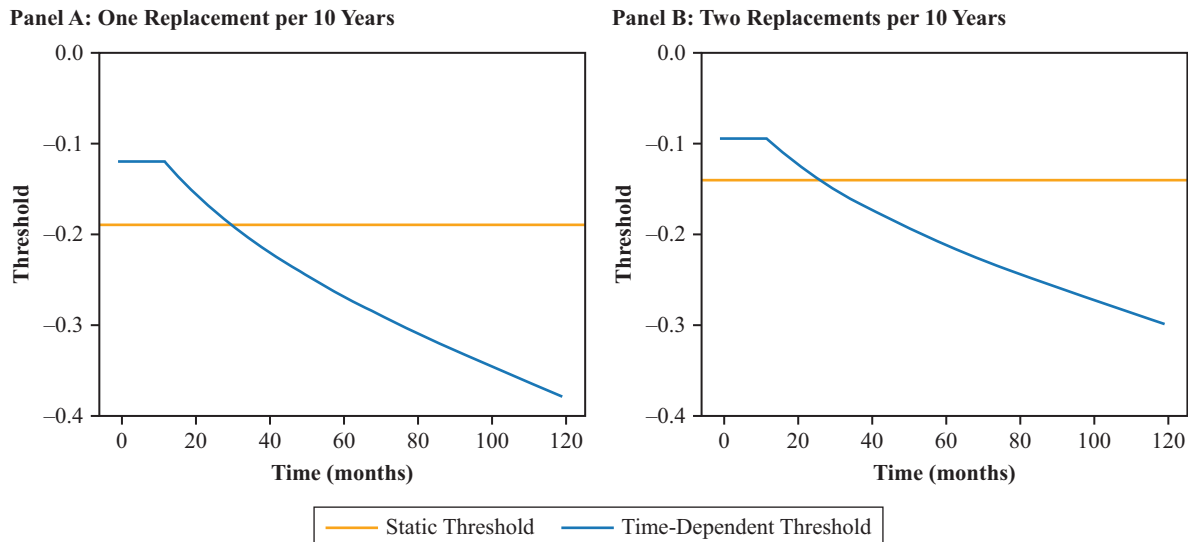
In Exhibit 11, we illustrate the effect of a drawdown-based rule, in which a risk reduction of 50% is triggered if the drawdown dips below a cutoff value. Full risk taking is restored if the manager recoups half of the losses (i.e., the manager would have recovered the peak-to-trough loss if risk had 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. Although the Sharpe ratio can improve slightly from such a risk reduction rule, the annualized return is lower. This is perhaps an obvious result because there is always still a chance the manager is *good* (0.5 Sharpe ratio). In the worst case that the manager is *bad*, it has a zero (and not negative) Sharpe ratio.

Although 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, 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 manager with a negative Sharpe ratio, but it seems a stretch because it requires negative skill or very high transaction costs. An alternative is a setting with a very high degree of autocorrelation, in which one may

Exhibit 10

Equivalent Thresholds through Time for the Different Drawdown Replacement Rules

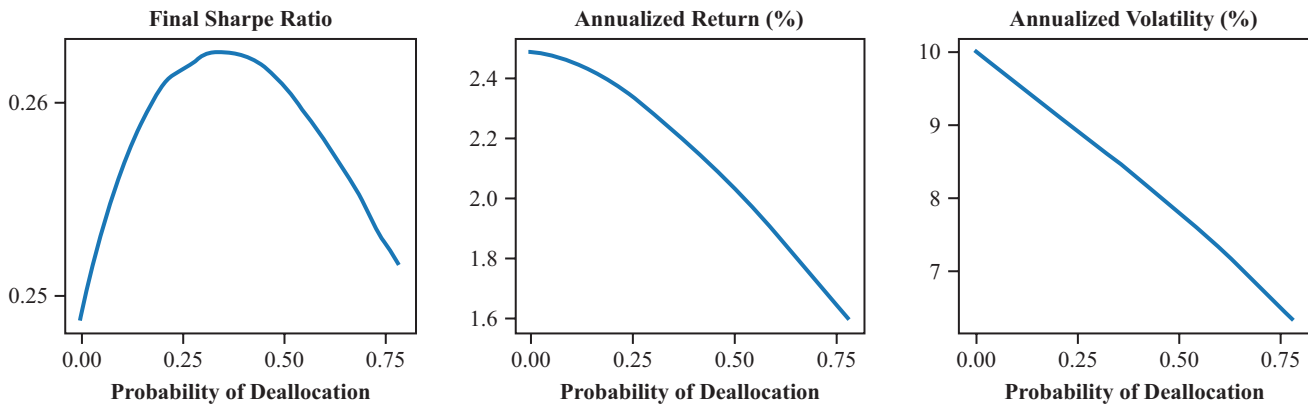


Notes: Equivalent threshold through time used for each drawdown rule for one and two replacements over 10 years. This figure illustrates the thresholds used to create Exhibit 9. Panel A shows the equivalent threshold used through time for one replacement per 10-year period, and Panel B shows the equivalent threshold for two replacements per 10-year period.

Source: Man AHL.

Exhibit 11

Efficacy of Risk Reduction Rule, Monthly Evaluation, No Type Migration



Notes: 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). Monthly returns are normal, i.i.d., with 10% annualized volatility for both manager types. We vary the cutoff value used in the risk reduction rule and plot the average Sharpe ratio, annualized return, and annualized volatility against the probability of deallocation.

Source: Man AHL.

have a negative expected return following a negative realized return.¹⁴

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 article.

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 preset drawdown rule may prevent peak risk taking. Taking risk in bursts (leading to heteroskedastic 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. Furthermore, automatic deallocation at a given drawdown level may reduce the chance of adverse behavior to exploit the trader put (i.e., manager increasing risk when returns fall to maximize chances of recovery), whereas 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 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 costly. However, if bad managers have negative Sharpe ratios (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 in which a manager abruptly loses skill. In reality, the two are complementary; the relative weight on total returns versus drawdown depends 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 (e.g., 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.

¹⁴It 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 ratio of the process; see Kaminski and Lo (2014).

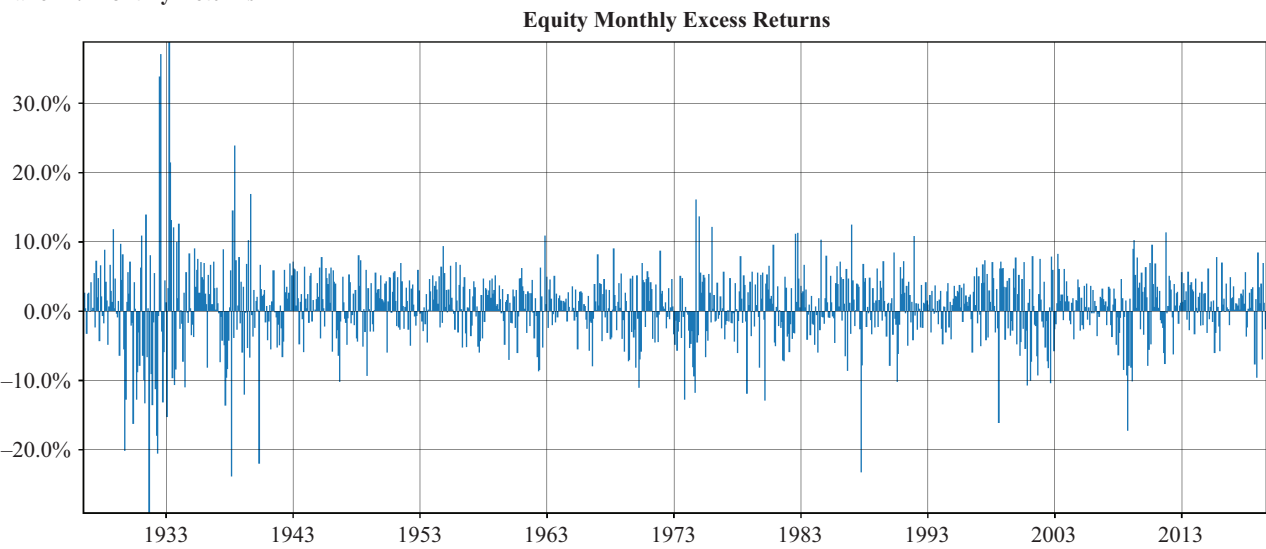
Appendix A

HETEROSKEDASTICITY FOR US STOCKS

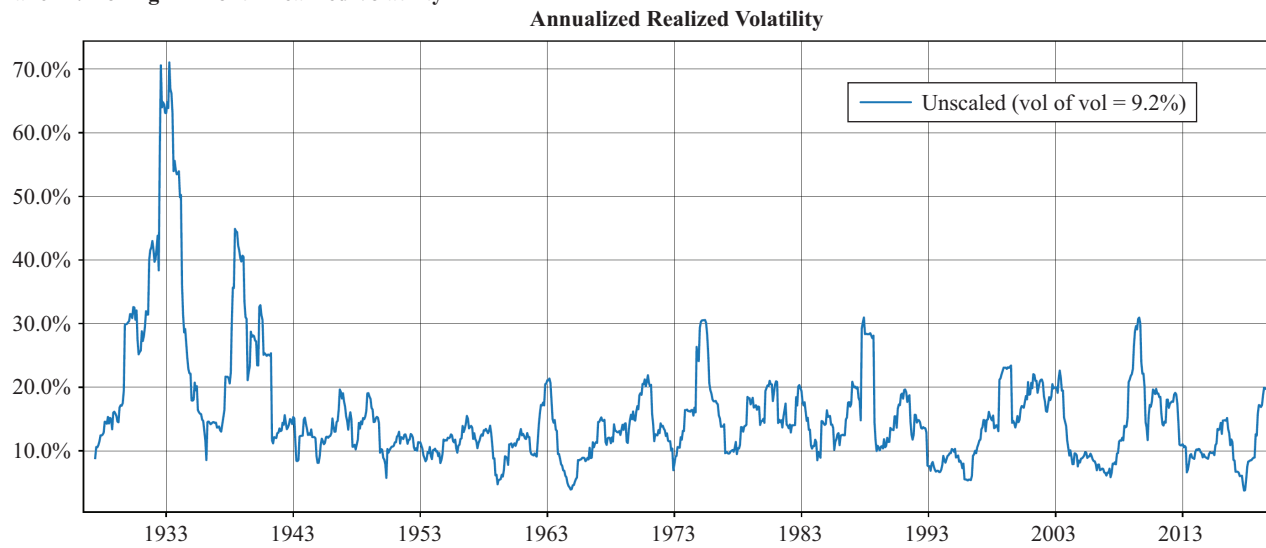
In Exhibit A1, we show monthly US equity returns (Panel A) and the rolling 12-month realized volatility for US equity returns from June 1926 to December 2019. Volatility is persistently high around, for example, 1929 (Great Depression) and 2008 (Global Financial Crisis).

Exhibit A1 US Equity Market Performance

Panel A: Monthly Returns



Panel B: Rolling 12-Month Realized Volatility



Note: Monthly US equity returns (Panel A) and the rolling 12-month realized volatility for US equity returns from June 1926 to December 2019.

Source: Ken French's data library (mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

Appendix B

HISTORICAL GAP MOVES

In Exhibit B1, we list the most negative monthly returns for a range of securities.

Exhibit B1

Historical Biggest Negative (Gap) Moves

	Sample Period		Most Negative Return			Vol of Vol		
Description	Start	End	Month	Value	#STDs (ann)	(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	

Note: Number of liquid securities the sample period (start date, end date, total number of years) and the worst monthly return (which month, percentage return, and number of annual standard deviations).

Source: Man AHL and Bloomberg.

ACKNOWLEDGMENTS

The authors would like to thank Nick Barberis, Anthony Ledford, Matthew Sargaison, and Rikiya Takahashi for valuable comments.

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." *The 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." *International 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." *The 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." *The Journal of Portfolio Management* 45 (1): 14–33.
- Harvey, C. R., and Y. Liu. 2020. "False (and Missed) Discoveries in Financial Economics." *The 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." *The Journal of Alternative Investments* 7 (4): 83–91.
- Magdon-Ismael, M., A. Atiya, A. Pratap, and Y. Abu-Mostafa. 2004. "On the Maximum Drawdown of a Brownian Motion." *Journal of Applied Probability* 41 (1): 147–161.
- Molyboga, M., and C. L'Ahelec. 2017. "Portfolio Management with Drawdown-Based Measures." *The Journal of Alternative Investments* 19 (3): 75–89.
- Sornette, D. *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton: Princeton University Press. 2003.
- 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.

To order reprints of this article, please contact David Rowe at d.rowe@pageantmedia.com or 646-891-2157.

ADDITIONAL READING

Portfolio Management with Drawdown-Based Measures

MARAT MOLYBOGA AND CHRISTOPHE L'ACHELEC

The Journal of Alternative Investments

<https://jai.pm-research.com/content/19/3/75>

ABSTRACT: This article analyzes the portfolio management implications of using drawdown-based measures in allocation decisions. The authors introduce modified conditional expected drawdown (MCED), a new risk measure that is derived from portfolio drawdowns, or peak-to-trough losses, of demeaned constituents. They show that MCED exhibits the attractive properties of coherent risk measures that are present in conditional expected drawdown (CED) but are lacking in the historical maximum drawdown (MDD) commonly used in the industry. This article introduces a robust block bootstrap approach to calculating CED, MCED, and marginal contributions from portfolio constituents. First, the authors show that MCED is less sensitive to sample error than CED and MDD. Second, they evaluate several drawdown-based minimum risk and equal-risk allocation approaches within the large-scale simulation framework of Molyboga and L'Ahelec via a subset of hedge funds in the managed futures space that contains 613 live and 1,384 defunct funds over the 1993–2015 period. The authors find that the MCED-based equal-risk approach dominates the other drawdown-based techniques but does not consistently outperform the simple equal volatility-adjusted approach. This finding highlights the importance of carefully accounting for sample error, as reported by DeMiquel et al., and cautions against overreliance on drawdown-based measures in portfolio management.

Maximum Drawdown

Models and Applications

RICARDO PEREIRA CÂMARA LEAL AND BEATRIZ VAZ DE MELO MENDES

The Journal of Alternative Investments

<https://jai.pm-research.com/content/7/4/83>

ABSTRACT: A drawdown is defined as the accumulated percentage loss due to a sequence of drops in the price of an investment. It is collected over non-fixed time intervals, and its duration is also a random variable. The maximum drawdown occurring over a fixed investment horizon is a flexible measure that provides a different perception of the risk and price flow of this investment. In this article,

a new modeling strategy for maximum drawdown that separates its duration and severity along with suggestions of a flexible distribution from the extreme value theory for modeling the losses is proposed. Applications for the fitted distribution, which include the computation of risk measures, classification and selection of investments, as well as its use in optimal portfolio allocation, are discussed. The properties of the model-based maximum drawdown-at-risk are then analyzed and a new risk measure, the conditional maximum drawdown-at-risk, is presented. Illustrations using stock market indexes as well as insights on the connection of the fluctuations of the maximum drawdown and the volatility of the underlying return series are provided. This methodology is applicable to various alternative investment strategies as well.

The Impact of Volatility Targeting

CAMPBELL R. HARVEY, EDWARD HOYLE, RUSSELL

KORGAONKAR, SANDY RATTRAY, MATTHEW

SARGAISON, AND OTTO VAN HEMERT

The Journal of Portfolio Management

<https://jpm.pm-research.com/content/45/1/14>

ABSTRACT: Recent studies show that volatility-managed equity portfolios realize higher Sharpe ratios than portfolios with a constant notional exposure. The authors show that this result only holds for risk assets, such as equity and credit, and they link this finding to the so-called leverage effect for those assets. In contrast, for bonds, currencies, and commodities, the impact of volatility targeting on the Sharpe ratio is negligible. However, the impact of volatility targeting goes beyond the Sharpe ratio: It reduces the likelihood of extreme returns across all asset classes. Particularly relevant for investors, left-tail events tend to be less severe because they typically occur at times of elevated volatility, when a target-volatility portfolio has a relatively small notional exposure. We also consider the popular 60–40 equity–bond balanced portfolio and an equity–bond–credit–commodity risk parity portfolio. Volatility scaling at both the asset and portfolio level improves Sharpe ratios and reduces the likelihood of tail events.