THE PROS AND CONS OF "DRAWDOWN" AS A STATISTICAL MEASURE OF RISK FOR INVESTMENTS

By David Harding, Georgia Nakou and Ali Nejjar, Winton Capital Management

A key measure of track record quality and strategy "riskiness" in the managed futures industry is drawdown, which measures the decline in net asset value from the historic high point. In this article we look at its strengths and weaknesses as a summary statistic, and examine some of its frequently overlooked features.

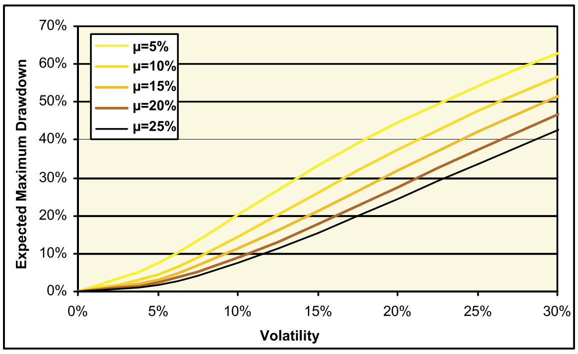
Under the Commodity Futures Trading Commission's (CFTC) mandatory disclosure regime, managed futures advisors are obliged to disclose as part of their capsule performance record their "worst peak-to-valley drawdown"[1]. As a description of an aspect of historical performance, drawdown has one key positive attribute: it refers to a physical reality, and as such it is less abstract than concepts such as volatility. It represents the amount by which you are less well off than you were; or, put differently, it measures the magnitude of the loss an investor could have incurred by investing with the manager in the past. Managers are obliged to wear their worst historical drawdown like a scarlet letter for the rest of their lives. However, this number is less straightforwardly indicative of manager quality as is often assumed. The seeming solidity of the drawdown statistic dissipates under closer examination, due to a host of limitations which are rarely explored sufficiently when assessing its significance as a guide to the future performance of an investment.

We will begin by exploring what drawdown *does* relate to. At best, one could say that there is some relationship between drawdown and two more meaningful statistics, namely return and variability. It is worth exploring in some detail precisely how these relate to one another. An investment with a long-term positive expected return can be expected to "drift" upward through time. In the event of this positive expected return arising, as it often does, from a stochastic or partly stochastic process, this upward drift will contain some random variation that will frequently cause the investment's value to fall below a previously attained peak. The distance below a historical peak is a drawdown. We can predict that drawdowns will be smaller if either (a) the upward drift is steeper, or (b) the variability of the process is lower. Drawdown is thus a function of the mean and the variability of the return process. However, without knowing *what* function and without having some insight into the return generating process, we cannot possibly know what the magnitude of the drawdown relates to.

Raw drawdown figures therefore have little value as a descriptive statistic, and almost none as a predictive statistic.

If, for the sake of argument, we were to assume that the return generating process can be fully characterised by a mean return of μ and a standard deviation of σ (i.e. that it is a normal process), we can examine analytically the relationship between expected maximum drawdown D in time T, and μ and σ (refer Figure 1).

Figure 1: Expected Maximum Drawdown as a function of Volatility for a range of expected return targets (μ) (T = 10 years, f = monthly)



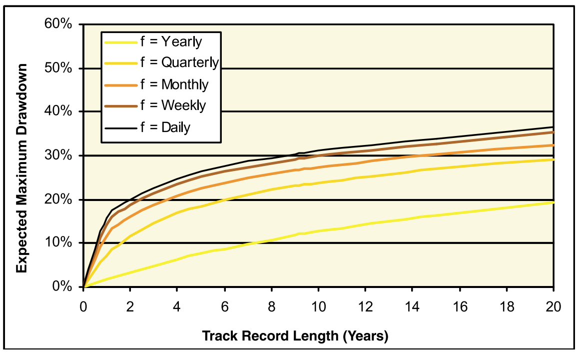
Subject to our comparing two investments with a similar σ over similar time periods, we can predict that a lower maximum drawdown would tend to be associated with a higher mean return μ, and thus a higher Sharpe ratio.

There is, however, a catch even to this simplified logic. Because maximum drawdown is a single number derived from a single string of data it is going to have a large error associated with it. This means that if we wish to extrapolate a future return or Sharpe ratio from a statistic employing maximum historical drawdown, this extrapolation will be highly error-prone and thus not necessarily a very useful predictor. As a result, comparing two processes of equal volatility, there is a considerable likelihood that the one with the higher expected return will have a greater maximum drawdown (the probability is exactly 50% if the expected returns are equal). Building statistical inferences on a single highly error-prone statistic is as secure as trying to balance a pencil on its tip! Thus, even if careful adjustment is made to equalise the volatility of track records, maximum drawdown is a poor statistic for making inferences about future reward/risk ratio or even future drawdown. Errors in statistical measures are usually reduced by averaging; thus, the average of the 10 worst five-day losses would be a much less error-prone statistic.

But that is not all. There are two further adjustments that need to be made in order for the maximum drawdown figure to have even the limited utility alluded to above.

The first is that, all other things being equal, drawdowns will be greater the greater the frequency of the measurement interval. The

Figure 2: Expected Maximum Drawdown as a function of Track Record Length for a range of measurement intervals (f) (σ = 20%, μ = 20%)



1 Its precise definition in the CFTC's terms is "the greatest cumulative percentage decline in month-end net asset value due to losses sustained by any account during any period in which the initial month-end net asset value is not equalled or exceeded by a subsequent month-end net asset value."

maximum drawdown will be greater on a daily time series than on a weekly one, and weekly will be greater than monthly. Investments that are marked to market daily, such as managed futures, may thus appear at a disadvantage to less frequently valued investments (e.g. hedge funds). However, it is never appropriate to compare drawdowns between time series with different reporting intervals without making an appropriate correction.

The second is that the maximum drawdown will be greater for a longer time series, so that managers with longer track records will tend to have deeper maximum drawdowns (refer Figure 2). This effect would have perverse consequences if the raw maximum drawdown were used to measure quality across the board, as, in general, managers that have survived longer have given evidence of professional competence through overcoming such adversities.

Thus, in order to make drawdown a more informative statistic, we must correct for track record length, measurement interval and volatility; we must take account of the error as well as making sure that we understand the nature of the return generating process (i.e. that it is reasonably parametric). Though some analysts correct for some of these factors, the conventional cursory use of drawdown as a statistic fails most or all of these tests, making it worse than useless. Rather than seeking to correct for all of these caveats, we may be better advised to focus directly on forecasts of return and variability. Drawdown may have a role in manager risk control, but it should be used with caution, and should be calculated with reference to probability (95%, 99% confidence level) from the characteristics of the underlying process rather than purely from the historical track record.