research *note*

CARR FUTURES
www.carrfutures.com

CARR FUTURES

SEPTEMBER 4, 2003

Understanding drawdowns

GALEN BURGHARDT • RYAN DUNCAN • LIANYAN LIU

Within the universe of hedge funds and commodity trading advisors (CTAs), one of the most widely quoted measures of risk is peak to trough drawdown. Our experience suggests, however, that investors do not have a widely accepted way of forming expectations about just how much managers who are in business over long periods of time might be expected to lose. Rather, we find that investors tend to monitor a manger's worst or maximum drawdown with only informal or anecdotal information about the manager's average annual or previous year's returns. Drawdown as a measure of risk has failed to attract the same kind of research and attention that are devoted to other common measures such as return volatility, VaR, or Sharpe ratios.

Our purpose here is to show that it is possible to get a reasonable fix on what drawdown distributions should look like. This is no trivial problem. Any manager for whom the standard deviation of returns is large enough to produce a loss in any given investment period will experience drawdowns. Most managers are in drawdown most of the time. And managers who have been in business a long time may well have experienced more and bigger drawdowns than those with short track records.

If it is possible to predict how drawdowns should behave, however, then we can address two important kinds of questions.

> ***Looking back*** over a manger's track record, does his drawdown history make sense? That is, do the frequency and size of his drawdowns look reasonable, and does his maximum drawdown accord with what we would expect.
>
> ***Looking forward***, what kinds of drawdowns should we expect over any given investment horizon. How many drawdowns should he experience? How big? How likely is it that his largest drawdown going forward will be greater than his maximum drawdown so far. And if it is bigger, how much bigger.

What we show here is that the three most important determinants of drawdowns are length of track record, mean return, and volatility of returns. The acid test, we think, is our simulated drawdown distributions do a very good job of explaining the kinds of drawdown patterns that CTAs have exhibited over the past 10 years.

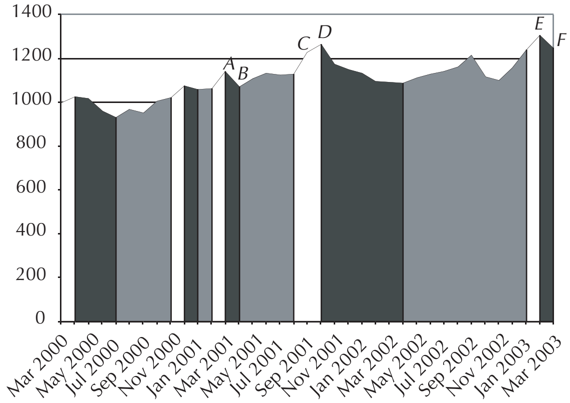
Drawdown defined

Drawdown measures the change in the value of a portfolio from any newly established peak (or high water mark) to a subsequent trough (or low water mark).

One of the things that makes drawdowns interesting is that they depend so much on the sequence of a manager's returns. The usual summary statistics such as mean and volatility of returns reflect nothing of the sequence in which the returns occur. The sequence is critical, however, for drawdowns. Two managers with identical means and volatilities of returns can experience very different drawdowns.

In practice, a drawdown is defined as the percent change in a manager's net asset value

Exhibit 1
Net asset value history with sample high and low water marks



from a high water mark to the next low water mark. A net asset value qualifies as a high water mark if it is higher than any previous net asset value and if it is followed by a loss. Thus, points A and D in Exhibit 1 are high water marks, but point C is not, even though net asset value at that point is higher than it has ever been. Point C is not a high water mark because it is followed by a gain.

A net asset value qualifies as a low water mark if it is the lowest net asset value between two high water marks. Point B qualifies as a low water mark. Or, if one is at the end of a data series, a low water mark is simply the lowest net asset value following the last high water mark. For example, point F, which follows the newly established high water mark at E, would be a low water mark for the purpose of calculating drawdowns even though the manager's net asset value has not yet reached a new high water mark.

A manager's maximum drawdown is simply the largest of these drawdowns.

What should drawdowns looks like?

Because realized drawdowns are the result of sequences of returns and depend entirely on the paths that a manager's net asset value can follow, the only practical way to discover what drawdowns should look like is to simulate as many net asset value paths as one needs to produce reasonable looking distributions. In what follows, we have used Monte Carlo simulations in which we have controlled for length of track record, the distribution of returns, deleveraging when in drawdown, and survival. The resulting drawdown distributions have two basic shapes.

For a given return distribution and length of track record, the frequency and size of a manager's entire collection of drawdowns will look like the distribution shown in Exhibit 2. In this exhibit (and in all of our drawdown exhibits), we show drawdowns as negative percent changes, and so we see in Exhibit 2 a high frequency of small drawdowns and a small frequency of large drawdowns.

Also, even though any given manager can only have one worst drawdown, it still makes sense to think of the distribution from which that worst or maximum drawdown was drawn. Or, if we think of several managers, all of whom have the same or very similar track records and return characteristics, we can think about what the distribution of their various worst drawdowns should look like. An example of what the distribution of maximum drawdowns should look like is provided in Exhibit 3.

Exhibit 2
The distribution of all drawdowns

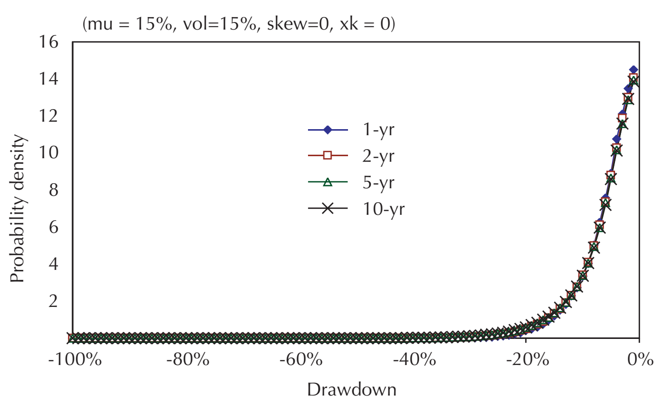
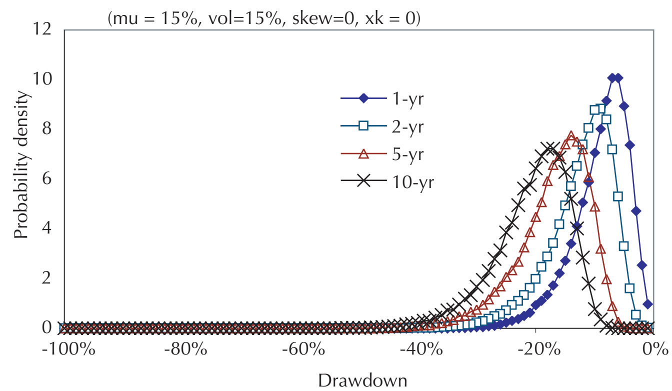


Exhibit 3
The distribution of maximum drawdowns



What forces shape drawdown distributions?

In our simulations, we were able to control the return generating process for most of the things that would seem to make sense. In particular, we controlled for:

- length of track record
- mean return
- volatility of returns
- skewness
- kurtosis
- deleveraging when in drawdown

Of these, the only three that have any empirical importance seem to be length of track record, mean return, and the volatility of returns. The rest tend not to matter much, in some cases because the effect of a change in the variable is small and in others because the range of the variable is small.

The distribution of all drawdowns

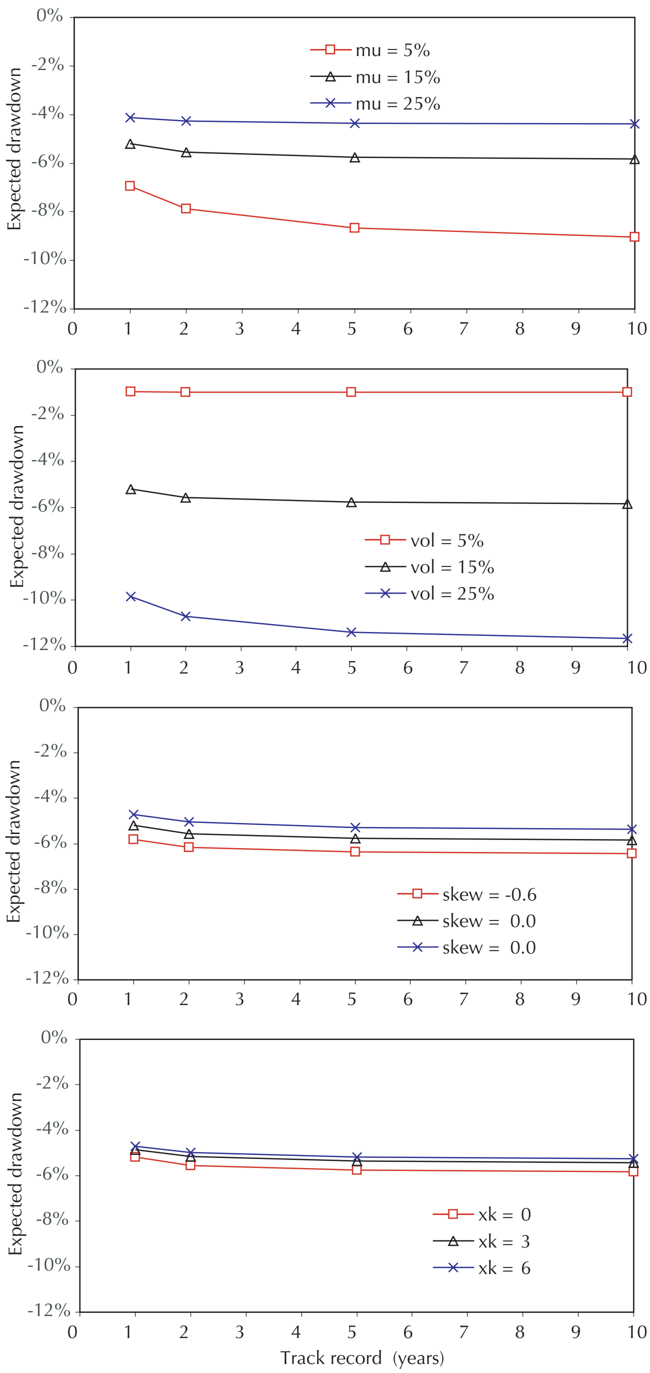
From Exhibit 2, we can see that length of track record matters very little to the distribution of all drawdowns. In other words, the likelihood of experiencing a drawdown of any given size is largely independent of how long a manager is in business.

Mean return and the volatility of returns, however, matter a lot. Exhibit 4 shows a manager's average or expected drawdown versus length of track record for different values of the four key moments of the return distribution. In the top panel, for example, we have varied the manager's mean return while holding volatility, skewness, and excess kurtosis (xk) constant. And, as one would expect, higher mean returns lead to smaller expected drawdowns.

From the second panel in Exhibit 4, it is also apparent that the volatility of returns has a large influence over a manager's drawdowns. Higher volatility leads to larger expected drawdowns.

Skewness and kurtosis, on the other hand, matter very little, at least given the range of values for skewness and kurtosis that we have observed in CTA returns over the past ten years. The most plausible reason for this seems to be that drawdowns are the result of adding together sequences of returns. As a result, even though the distribution from which any given return is drawn may be higher skewed or exhibit fat tails, the result of adding returns together produces a random variable that tends (a la the central limit theorem) to be more normally distributed.

Exhibit 4
Track record, mean return, and volatility of returns have the greatest effect on expected drawdowns



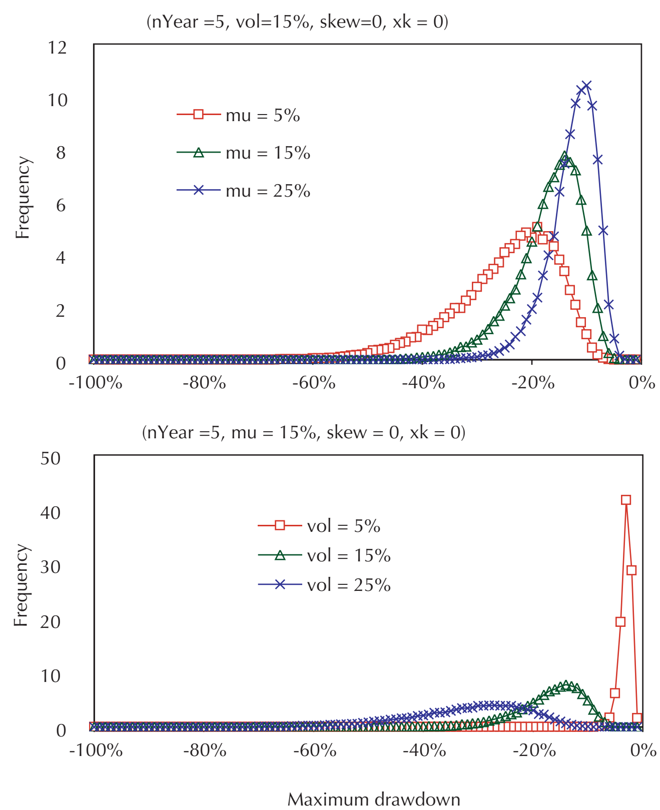
The distribution of maximum drawdowns

A manager's worst drawdown is taken from a distribution that is highly sensitive to length of track, mean return, and volatility of returns. Skewness and kurtosis, on the other hand, tend not to matter much.

The likelihood of any given drawdown is independent of how long a manager is in business. But the likelihood of experiencing a drawdown that is bigger than anything experienced so far increases with every passing day. As a result, as shown in Exhibit 3, increases in the length of track record shift the entire maximum drawdown distribution to the left.

Exhibit 5 shows how the distribution is affected by mean returns and the standard deviation of returns. The values we have chosen here correspond roughly to the range of values we observe in our data base of CTA

Exhibit 5
The effects of returns and volatility on maximum drawdowns



returns. As the upper panel shows, high returns tend to produce smaller maximum drawdowns, while the lower panel shows how increases in the volatility of returns increases the likelihood of large maximum drawdowns.

Exhibit 6 compares the effects of changing each of the four return characteristics on a manager's expected maximum drawdown. In each panel, it is apparent that length of track record matters more than it did with the expected value of all drawdowns. Each of these curves is considerably steeper here than those in Exhibit 4. It is also apparent that mean return and volatility of returns matter a lot, while skewness and kurtosis matter hardly at all.

To put the importance of these things in perspective, we calculated the partial effect of each moment on expected maximum drawdown and multiplied the partials by the standard deviation of each moment as measured in our database. The results are shown in Exhibit 7, which shows that volatility of returns is by far and away the most important variable across managers. Variation in mean returns is a strong second. In contrast, skewness and kurtosis rank very low on the list of things that matter for an explanation of why different managers have different drawdowns.

Exhibit 6
The three most important variables for maximum drawdown are length of track record, mean return, and volatility

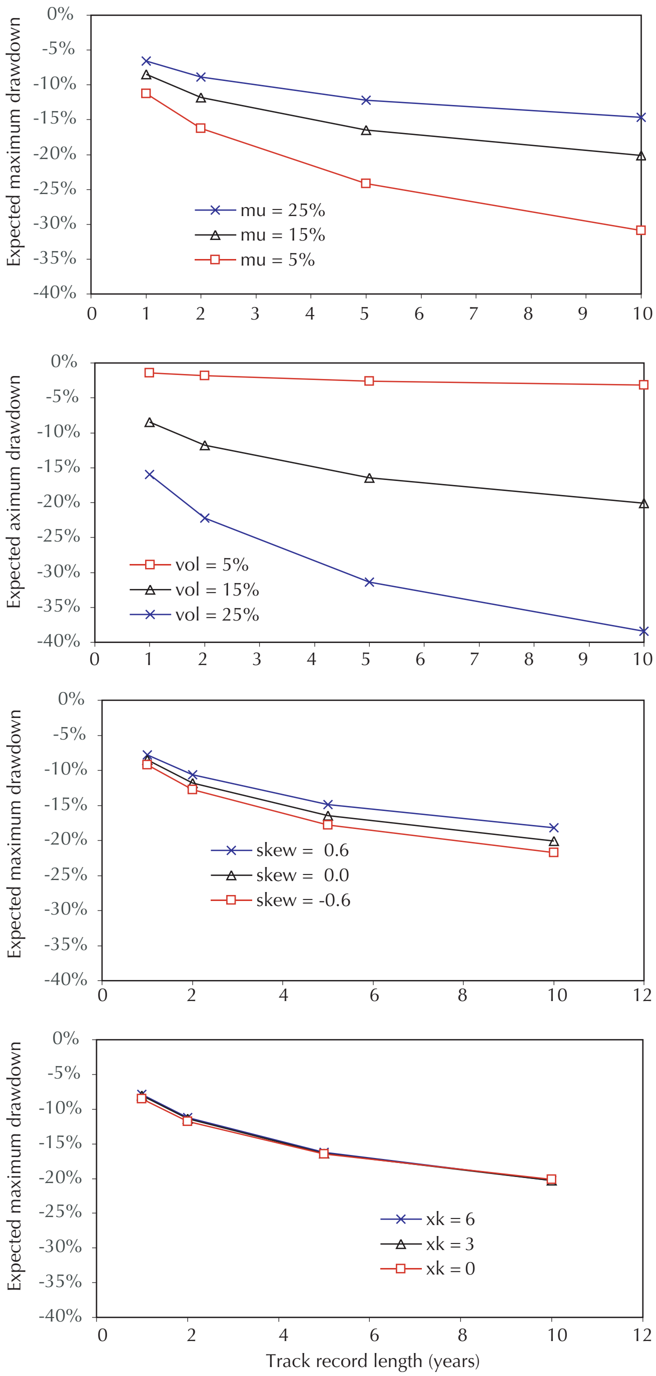


Exhibit 7
How important are the four moments for expected maximum drawdowns?

Moment	Potential effect: partial effect* (%)	standard deviation of moment	combined effect (%)
Mean	68.0	0.139	9.5
Standard deviation	-191.0	0.133	-25.4
Skewness	2.7	1.035	2.8
Kurtosis	0.1	5.914	0.4

**local partial of expected maximum drawdown wrt each moment*

The core drawdown function

If higher returns produce smaller drawdowns while higher volatilities produce larger drawdowns, then one can trade off one for the other to produce the same expected drawdowns. But given the sizes of their respect effects, it can take a lot of extra return to make up for a little extra volatility.

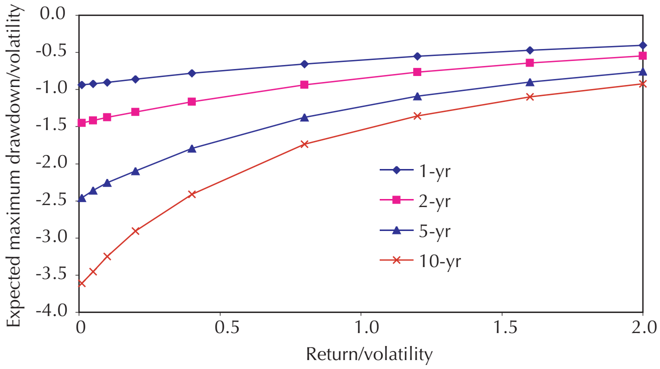
Even though there is no clean, analytical function that relates drawdowns to a manager's returns, the relationship between a manager's drawdowns and his returns and risks can be described by

$$DD/\sigma = f(\mu/\sigma)$$

where σ is the standard deviation of returns and μ is the mean return. That is, a manager's drawdowns, when divided by the volatility of returns, can be written as a function of the manager's modified Sharpe ratio (i.e., the ratio of mean return to the standard deviation of returns).

The shape of this function is illustrated in Exhibit 8 for track records ranging from 1 to 10 years. The curvature bears out our sense that volatility matters more than mean return. A doubling of a manager's mean return while holding return volatility constant will reduce expected drawdown per unit of volatility but by less than half. In turn, a doubling of volatility while holding mean return constant will more than double expected maximum drawdown per unit of volatility.

Exhibit 8
The shape of the relationship between expected maximum drawdowns and returns when both are normalized for volatility



If we are concerned only about the sizes of drawdowns, as opposed to drawdowns per unit of volatility, this relationship can be rewritten as

$$DD = \sigma f(\mu/\sigma)$$

which suggests the following:

a doubling of both mean return and volatility (which would leave the modified Sharpe ratio unchanged) will exactly double expected maximum drawdowns

a doubling of volatility alone will more than double expected maximum drawdowns

one would have to more than double mean return to compensate for a doubling of volatility

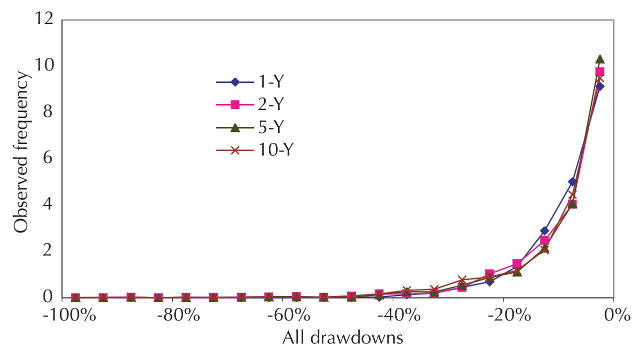
These points help to illustrate the differences between drawdown, volatility of returns, and a modified Sharpe ratio as measures of risk. All three are related but provide different perspectives. Two managers with the same volatility of returns will have different expected drawdowns if their mean returns are different. Two managers with identical modified Sharpe ratios will have different expected drawdowns if their return volatilities are different.

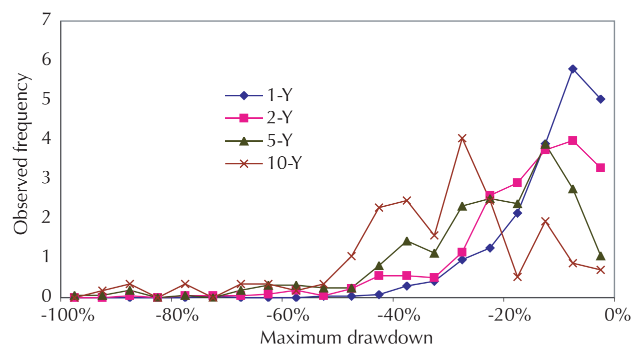
Empirical drawdown distributions

To see whether this approach could be used to explain the drawdown patterns we observe, we constructed drawdown histories for 1, 2, 5, and 10 years in the following way using CTA returns from the Barclays database. Using return histories for all managers with a 1-year track record as of November 2002, we determined what their drawdowns would have been had they all started from scratch at the end of November 2001. Then, for all managers who had a 2-year track record as of November 2002, we determined what their drawdowns would have been had they all started fresh at the end of November 2000. And so forth for 5-year and 10-year track records. By design, this approach produces different drawdown histories than those actually reported by the CTAs in our database. It has the advantage, however, of putting all managers up against the same market conditions.

The results of these efforts are shown in Exhibit 9. The distributions of all drawdowns are shown in the upper panel, while the distributions of maximum drawdowns are shown in the lower panel.

Exhibit 9
Observed drawdown distributions
(1993 through 2002)





Reconciling theoretical and empirical distributions

The distribution of all drawdowns shown in Exhibit 9 look about the way we would expect. The distributions of maximum drawdowns, on the other hand, posed a real challenge. First, they are irregularly shaped. Second, as shown in Exhibit 10, where we focus on the distribution of maximum drawdowns for 10-year track records, the observed distribution does not fit well with the theoretical distribution (labeled "discrete") derived from the actual distribution of returns for all CTAs with a 10-year track record.

The problem is that the observed drawdown distribution peaks at a much lower level of drawdowns than does the theoretical. One plausible explanation for the difference in the shapes of the two distributions is that managers may deleverage when they are in drawdown

– that is, scale back the risk they take – and thereby avoid the larger drawdowns they would experience if they were to keep the volatility of returns constant.

Exhibit 10 shows that it is possible to pull the theoretical drawdown distribution to the right by allowing for deleveraging when simulating returns. In this case, we scaled the manager's mean and volatility of returns as

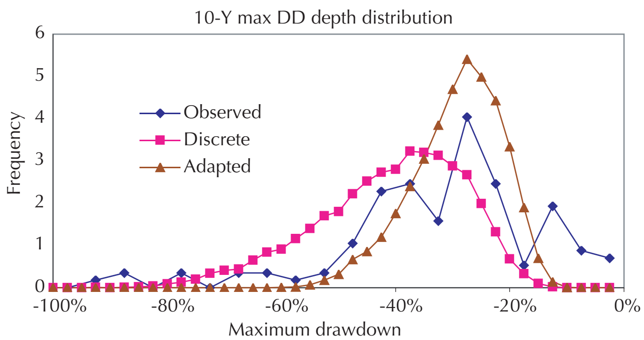
$$\mu' = f x \mu$$
$$\sigma' = f x \sigma$$

where

$$f = 1 - [\text{abs(drawdown)}]^{1/2}$$

so that if a manager's current drawdown were 50%, the scaling factor would be .29 [= 1 - $.5^{1/2}$]. The new distribution of maximum drawdowns that results is labeled "Adapted" in Exhibit 10 and peaks just about where it should.

Exhibit 10
First cut at explaining observed maximum drawdowns

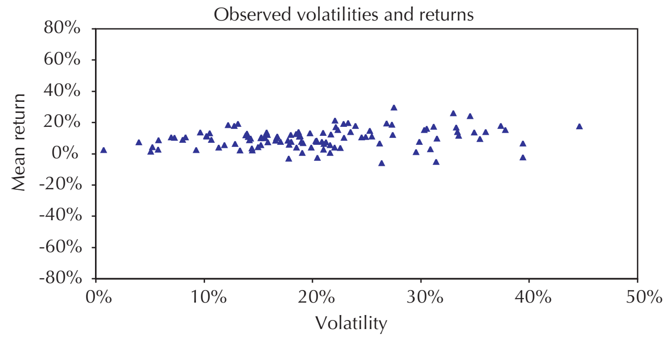


One main drawback to this approach is that the evidence on deleveraging is largely anecdotal. We know managers who attest to the fact that they scale back risk when in drawdown. We know other mangers, however, who say that they do not. And we have not been able yet to find any evidence in the volatilities of managers' returns that suggests that they deleverage when in drawdown.

Another main drawback to this approach is that while it produces a mass of probabilities that looks like what we observe, it greatly underpredicts the several large drawdowns that we observe in the data.

A better solution seems to lie in the fact that managers exhibit very different volatilities of returns. This is borne out in Exhibit 11, which shows a scatter plot of mean returns and their corresponding volatilities for those managers for whom we have 10 years of performance data. For the purposes of this exercise, we grouped the managers into three broad volatility groups: low (0% to 12.5%), medium (12.5% to 25%), and high (25% to 50%).

Exhibit 11
The range of returns and volatilities for managers with a 10-year track record



Managers grouped by volatility

Volatility range	Number of managers	Group return	Group volatility
0.0% ~ 12.5%	19	8.3%	8.6%
12.5% ~ 25.0%	64	9.4%	18.8%
25.0% ~ 50.0%	9	12.3%	32.4%

Using the group returns and group volatilities, we simulated the three maximum drawdown distribution shown in the upper panel of Exhibit 12. Then, using the numbers of managers in each of the three groups, we produced a composite distribution that is a weighted average of the three separate distributions.

The resulting composite distribution has some attractive features. First, because of the presence of the low volatility group, the composite distribution peaks about where it should. Second, because of the presence of the high volatility group, the composite distribution allows for a sufficiently high probability of large drawdowns. And third, as shown in the lower panel of Exhibit 12, the composite even exhibits some of the irregular shape that we see in the observed distribution of drawdowns.

Putting a manager's drawdown experience in perspective

At this point, we think it is reasonable to draw two theoretical drawdown distributions for any given manager, both based on length of track record and the mean and volatility of returns. We have done this in Exhibit 13 for a manager with a 10-year track record, a mean return of just over 12%, and a standard deviation of returns of 20%. Over this, we have superimposed the manager's actual drawdowns, which are represented by the horizontal lines stemming from the vertical axis on the right. This particular manager has experienced 17 drawdowns over the 10 years, most of them less than 10%. The maximum drawdown was just over 40%.

Overall, this manager's actual drawdown experience is roughly in line with what we would expect. The maximum drawdown is in the upper end of the theoreti-

Exhibit 12
A composite maximum drawdown distribution

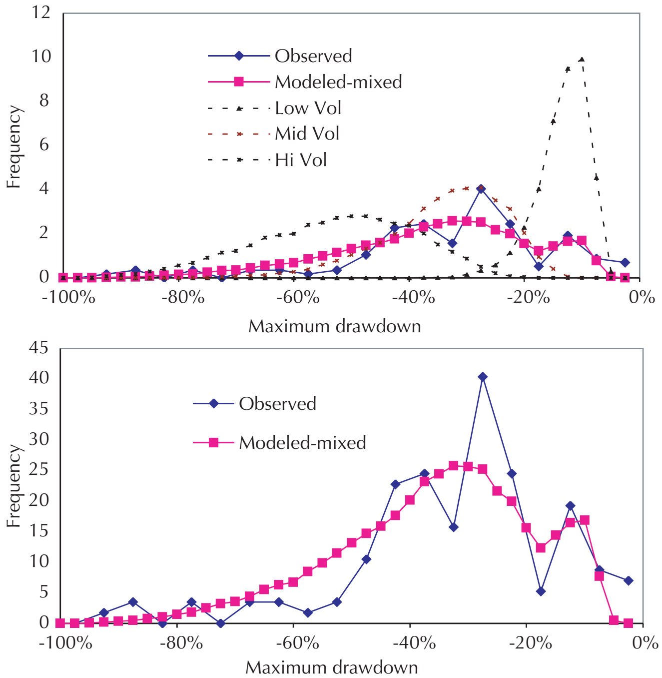
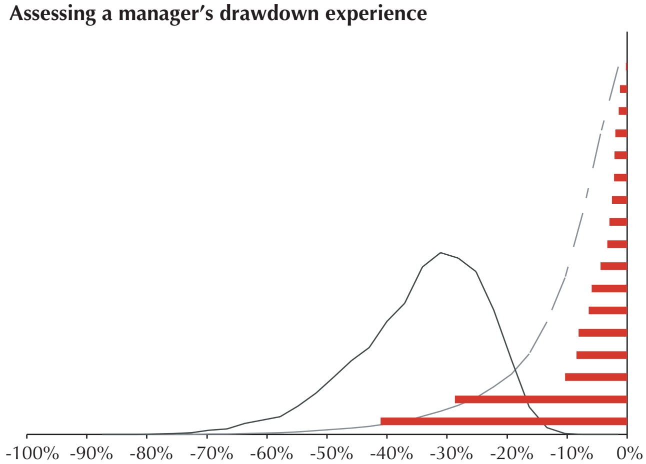


Exhibit 13
Assessing a manager's drawdown experience



cal distribution, but appears to be only about 1 standard deviation above the mean.

What about future drawdowns?

What kinds of drawdowns might an investor expect going forward? This work suggests that we can form reasonable expectations about the size and frequency of drawdowns for any given investment horizon. We can also say something useful about the possibility that a manager will experience a larger drawdown than the maximum drawdown to date. In particular, for a given investment horizon and assumptions about the mean and volatility of returns, we can calculate the likelihood that a manger will experience a worse drawdown and a conditional maximum drawdown to go with it.

Exhibit 14
How much worse can it get?

(mean return = 12%, volatility = 20%)

Investment horizon (years)	Probability of a drawdown > 41%	Expected drawdown if > 41%
1	0.1%	-44.1%
2	1.7%	-45.8%
3	4.3%	-47.1%
4	6.8%	-47.3%
5	9.5%	-47.8%
6	12.9%	-48.2%
7	15.3%	-48.4%
8	17.8%	-48.8%
9	20.5%	-49.1%
10	23.4%	-49.0%

For example, how likely is it that a manager whose worst drawdown to date is 41% will have a worse drawdown over any given investment horizon? If we are willing to assume a mean return and volatility (e.g., 12% and 20%), we find in Exhibit 14 that the probability of experiencing a drawdown greater than 41% is only .1% over the next year but would be 23.4% if the investment horizon is extended to 10 years. We also find that the expected value of this worse drawdown would be 44.1% if it occurs in the next year but would be 49.0% if experienced over a 10 year horizon.

In practice, we can use any target drawdown, not just the worst or maximum drawdown to date. And we can, if need be, modify the assumptions about the manager's returns to produce more realistic theoretical distributions. This would be especially useful if we think that a manager's trading strategy is likely to produce a mix of volatilities over time. The drawdown distributions for high and low return volatilities have very different shapes and could produce very different probabilities of large losses than one would get using an assumption of constant volatility.

Further questions

Our conversations with clients and colleagues about this work have generated several questions that deserve a closer look. How important, for example, are serially correlated returns? How reliable are our estimates of return volatilities? Would the analysis be better if one had daily rather than monthly return data? Would this analysis work as well for hedge funds as it does for commodity trading advisors? What happens if a manager's return volatility changes in response to drawdowns?

Our preliminary on these suggests the following. First, serially correlated returns could have a measurable effect on drawdown distributions, but we have found no evidence of serial correlation in CTA returns. Volatility estimates based on monthly return data can be subject to very large statistical errors and would be much improved, at least in the case of CTAs, if we had daily return data. To the extent one can get reliable return and volatility information about hedge funds, the analysis should work well. It is much harder, though, to get the same quality information about hedge funds as one can get for CTAs. And, while we know that some managers deleverage when in drawdown, the evidence on CTAs as a class is ambiguous.

We invite comments and questions. Contact Galen Burghardt at 312-762-1140 or gburghardt@carrfutures.com

Technical note

To simulate net asset value series where skewness and kurtosis are zero, we draw sample returns from a lognormal return distribution. To capture skewness and kurtosis, we sample returns from a generalized lambda distribution. The values of skewness and excess kurtosis used in this note were roughly consistent with the range of values we observed for CTAs in our database. From the return series, we construct net asset value series. And from these, we derive the simulated drawdowns that are used to produce the theoretical drawdown distributions. A typical run usually requires 10,000 iterations to produce a smooth distribution.