Nov. 14, 2012

Galen Burghardt
galen.burghardt@newedge.com

Lianyan Liu
lianyan.liu@newedge.com

It's the autocorrelation, stupid

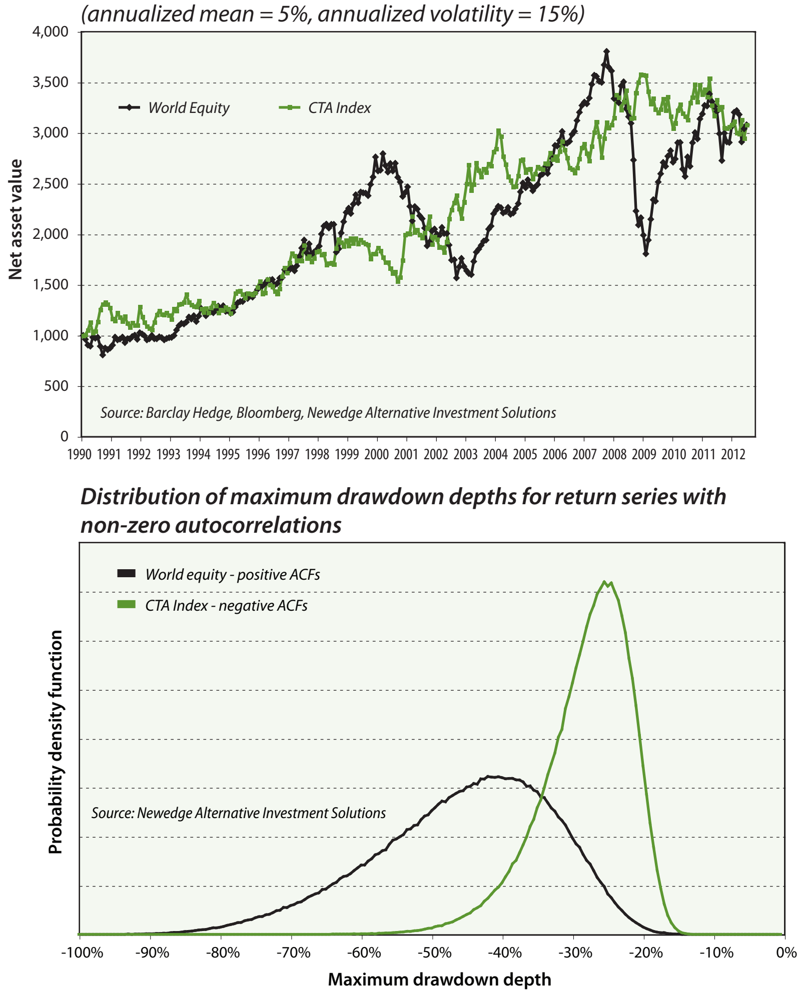
Drawdown – the money you've lost since reaching your most recent high water mark – could well be the single most discussed aspect of risk in the world of investing – especially in the world of active asset management. It represents money that once was within the investor's grasp but now is gone. The active manager can be in drawdown 80 percent of the time. It causes active managers to wonder if their approaches to trading are flawed. It causes investors to wonder if their active managers will change the way they trade in an effort to protect themselves. And, several years ago, when we polled the investors at one of our early gatherings about what they considered the most important measure of risk, drawdown topped the list. In fact, it was the results of this informal poll that led us to publish *Understanding drawdowns* in 2004, and the model we developed there has served us well for eight years.

Even so, the work of trying to understand drawdowns can reveal astonishing insights into the way we think about risk. In this case, we have learned that failing to account for autocorrelated returns can lead to serious biases in our estimates of return volatilities. Here is a case in point. While doing some work on pension fund investments, we raised the question about why drawdowns in equities can be so much deeper and last so much longer than one would expect given their volatility. In the upper panel of Exhibit 1, we have drawn two net asset value series – one for global equities and one for CTAs – in a way that imparts to each a mean return of 5% and a return volatility of 15% based on monthly returns. (See the appendix for a description of how these two series were derived.)

Exhibit 1

Net asset values for two return series

(annualized mean = 5%, annualized volatility = 15%)



It is clear to the eye, though, that these two series are very different and that equities exhibit much more risk. The difference, as we will show in this note, can be explained by the fact that equity returns tend to exhibit positive autocorrelation while CTA returns tend to exhibit negative autocorrelation. As a result, the standard square root of time rule that the industry uses to translate daily, weekly, or monthly volatilities into annualized volatilities is wrong and produced biased results.

Prime Clearing Services | Advisory Group

Galen Burghardt
galen.burghardt@newedge.com

Ryan Duncan
ryan.duncan@newedge.com

Alex Hill
alex.hill@newedge.com

Chris Kennedy
chris.kennedy@newedge.com

Lauren Lei
lauren.lei@newedge.com

Lianyan Liu
lianyan.liu@newedge.com

Jovita Miranda
jovita.miranda@newedge.com

James Skeggs
james.skeggs@newedge.com

Brian Walls
brian.walls@newedge.com

Tom Wrobel
tom.wrobel@newedge.com

newedge.com

The effects of autocorrelation can be huge. In the lower panel of Exhibit 1, we have drawn two maximum drawdown distributions. Underlying assumptions about length of track record, mean return and return volatility are the same in both cases, but the expected maximum drawdown with the positive autocorrelation shown by equities is nearly double that for an asset with the negative autocorrelation shown by CTAs.

The main conclusions of this round of work are these:

- CTA index returns in general, and the returns of trend following CTAs in particular, seem to exhibit negative autocorrelations that are both statistically significant, persistent through time, and persistent across trend following CTAs.
- For trend following CTAs, our failure to incorporate autocorrelation into our volatility measures has produced estimates that are much too high. The 15% return volatility for CTAs in Exhibit 1 really should be closer to 9.4%. (And the 15% return volatility for global equities should be closer to 17.9%.)
- By incorporating autocorrelation estimates into our drawdown model, we can produce much better estimates of the kinds of drawdowns we should expect for all liquid hedge fund strategies, not just trend following CTAs.
- As long as one does it with care, one can use autocorrelation estimates as an effective diagnostic tool to uncover influences on return volatilities that otherwise could take decades to find.

In the note that follows, we

- Describe the data sets that we used to establish the presence and persistence of autocorrelation in CTAs' returns
- Present a drawdown puzzle involving the 67 CTAs who had ever appeared in the Newedge CTA Index that brought the importance of autocorrelated returns to our attention
- Show how autocorrelated returns provide the key to unlocking the puzzle and the effect autocorrelation has on the way we should translate single-period volatilities into multi-period volatilities.
- Report on what we found when we examined other data sets of CTA returns.
- Return to the global equity/CTA comparison and conclude the note with an analysis of how autocorrelation affects risk and biases our measures of risk-adjusted returns in a potentially big way.

The data sets we used to establish the presence and persistence of autocorrelated returns

When we first presented our findings on autocorrelated returns and their importance for drawdowns and risk measures, the data we used were mainly monthly returns for the 67 CTAs who had ever appeared in the Newedge CTA Index. Because these are among the largest and most successful CTAs in the industry, this was a respectable data set. Most of the questions, though, focused on whether the results would hold up in the face of other data and on whether the results were stable through time. Here are descriptions of the data we've used.

A concatenated CTA index (January 1990 through July 2012)

This index – modified to produce a 5% mean return and a 15% return volatility – appears in Exhibit 1. To construct this index, we chained together the Barclay CTA index from 1990 through December 1999 and the Newedge CTA index from January 2000 through July 2012. While the Barclay CTA index comprises a much broader set of CTAs than does the Newedge CTA index, they correlate well with one another and are both free of most of the biases that affect indexes of self-reported returns.

CTAs that have appeared in the Newedge CTA Index

This was a natural data set for us to use. The index is based on returns of the 20 or so largest CTAs that are open for business and willing to provide daily return data. The index went live January 2000 and since then 67 different CTAs have been used in the calculation of the index. Because the number of CTAs is

Exhibit 2

Tallies for Newedge CTA Index constituents

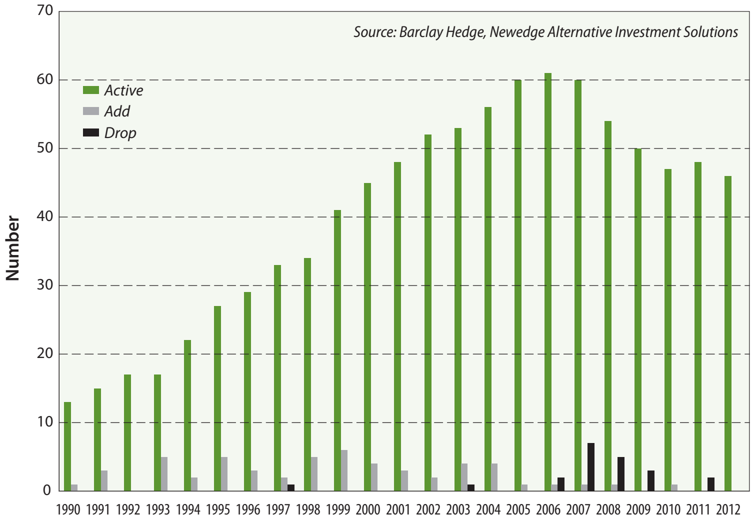
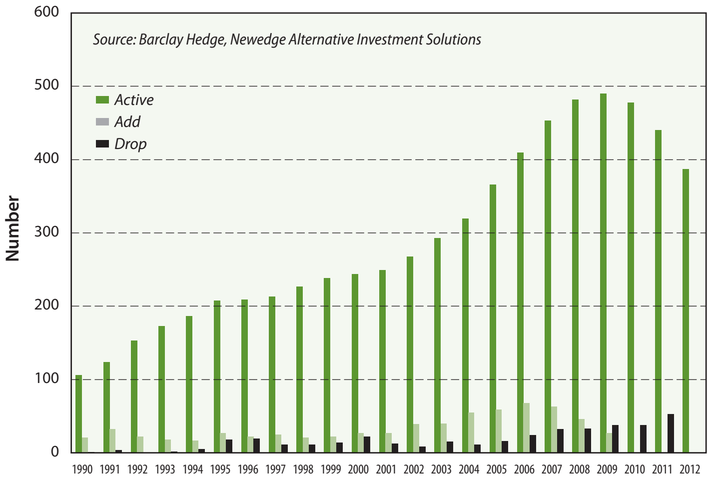


Exhibit 3

Tallies for larger CTA dataset



relatively small in this case and because we know them so well, it was easy for us to create two subsets – one for trend followers and one for non-trend followers. Our classification was based chiefly on reputation and confirmed correlation.

For each CTA, we used the full track record available to us beginning with January 1977. Exhibit 2 shows how the number of CTAs in this set reporting returns during any given year varied through time. As the exhibit shows, the number of CTAs increased more or less without interruption until 2006 and 2007. A history of adds and drops is provided as well.

All CTAs with $50 million and a 3-year track record

To determine whether what we found with the 67 CTAs in the previous data set would hold up in a broader data set, we pulled together a data set that comprised the returns of all CTAs who ever had $50 million under management and a three-year track record. A summary tally of the population of this set is provided in Exhibit 3. Altogether, the set includes the returns of 783 CTAs. A summary tally of how this data set evolved is provided in Exhibit 3.

A complete description of the work that went into assembling these data will be provided in a separate note. For now, it is enough to note that special care was taken to avoid duplications and to isolate what would be the lead program for each CTA.

A mini trend index

To mimic the approach we used to identify trend and non-trend managers in the Newedge CTA Index (ie. reputation and correlation), we needed to create a trend-following index with a much longer track record than that of the Newedge Trend Sub-Index. In doing so, we identified five well-known trend followers with continuous monthly returns from February 1988 through July 2012. From these five return series, we constructed an equally weighted index that we rebalanced at the end of each year.

While this index helped us to sort the larger data set based on correlation it was also used to check for persistence in autocorrelation though time.

The Newedge Trend Indicator

The last data set we have is not a CTA, but the Newedge Trend Indicator, which uses a 20/120 moving average trend following model to trade 55 markets that fall into four sectors – equities, currencies, interest rates and commodities. The construction of this index is described in Two Benchmarks for Momentum Trading.

Our findings

The focus of this section is on the importance of autocorrelated returns for the ways we think of risk and how we can use estimates of autocorrelation to improve our expectations for drawdowns and other measures of risk.

A second empirical puzzle

This work was inspired by two empirical problems. The first, we described in the introduction. That is, why could two return series – one for equities and one for CTAs – with identical means and volatilities produce such wildly different drawdown experiences.

Exhibit 4

Expected and observed maximum drawdowns for components of the Newedge CTA Index

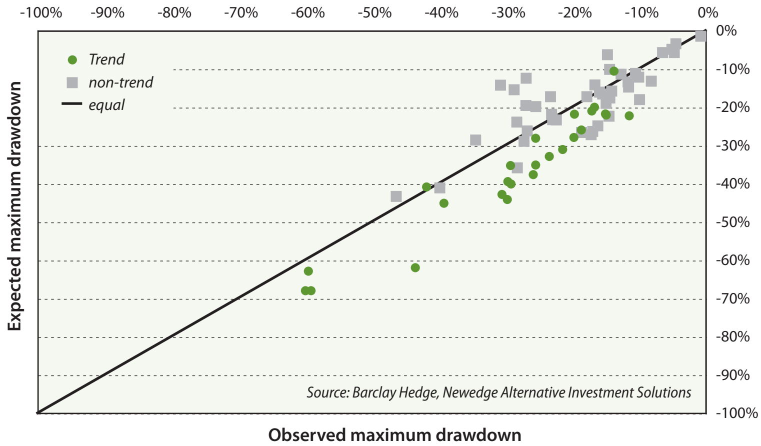


Exhibit 5

p-values of observed maximum drawdowns for CTA Index trend components

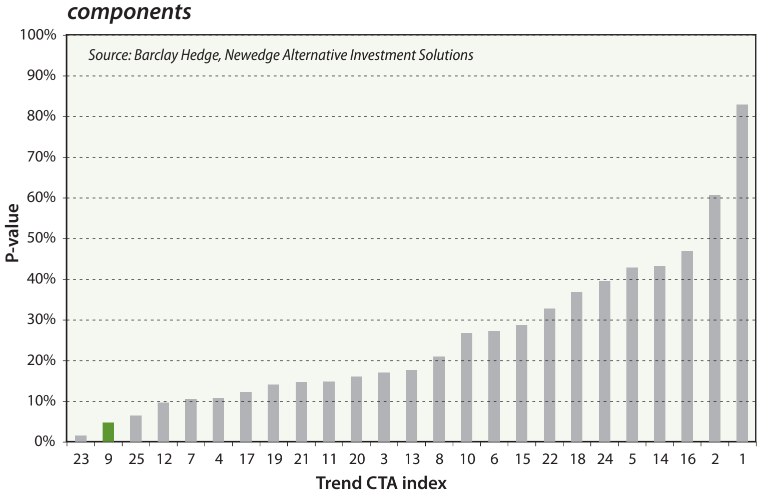
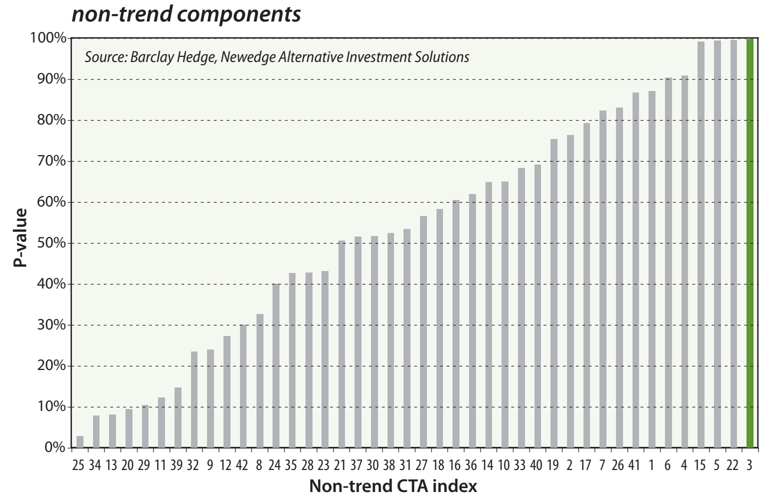


Exhibit 6

p-values of observed maximum drawdowns for CTA Index non-trend components



A second was what we found when we used our drawdown model to predict what various CTAs' maximum drawdowns should have been. For this exercise, we used all CTAs who had ever been part of the Newedge CTA Index anytime from 2000 through April 2012. In all, there were 67 CTAs of whom 25 would generally be recognized as trend followers. For each of these CTAs, we used the entire recorded track record irrespective of when it was included in the index. In each case, we calculated the CTA's mean return and return volatility and used the length of the CTA's track record to reckon his expected maximum drawdown.

What happened when we did this is shown in Exhibit 4 where we plot each CTA's expected maximum drawdown against the realized maximum drawdown. The upper right hand corner in this exhibit represents 0, and drawdowns are shown in negative numbers with expected drawdowns measured along the vertical axis and realized drawdowns measured along the horizontal axis. The line that runs from the lower left hand corner to the upper right hand corner shows where the two are equal. For any observation below the line, the realized maximum drawdown was smaller than the model predicted. For any observation above the line, the realized drawdown was larger than predicted.

The puzzle that leaps off the page most clearly is that with only two exceptions, the trend following CTAs' realized maximum drawdowns were smaller than the model predicted. One doesn't really need much more convincing that something is up, but in Exhibit 5, we show the p-values for the trend following CTAs. If our drawdown model were doing well, these p-values would follow a more or less straight line from the lower left hand corner to the upper right so that the values would be uniformly distributed from 0.00 to 1.00 with two or three CTAs' results falling in each 0.10 band. But what we see is that all but two of the values were less than 50% and most of these were less than 20%. (The numeric labels along the horizontal axis in this exhibit and the next represent the CTAs' places on our original list and bear no relationship to any ordering related to name, return, volatility, or age.)

Actually, a less obvious but related puzzle in-

Exhibit 7
Autocorrelations for trend CTA #9
(Sum of correlations = -0.42)

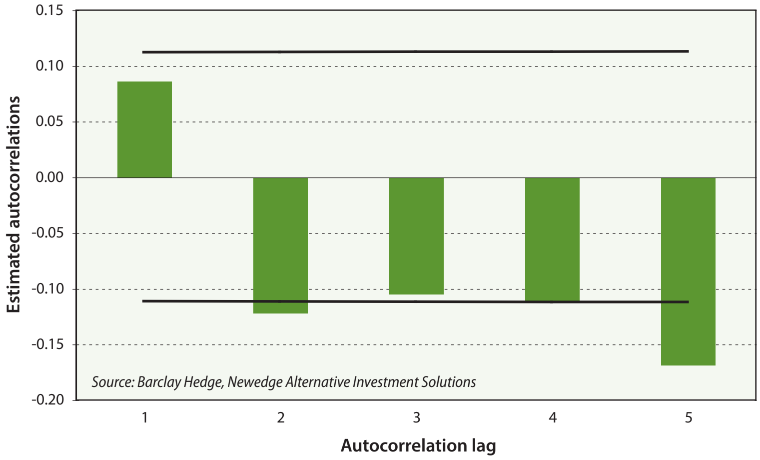


Exhibit 8
Autocorrelations for trend components

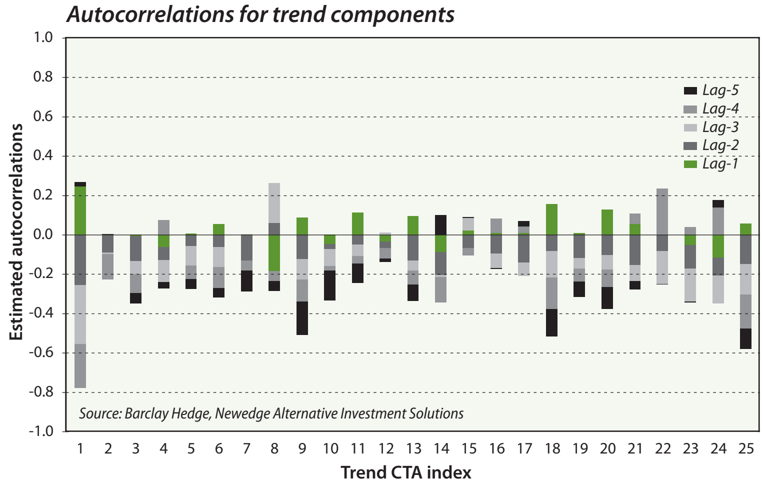
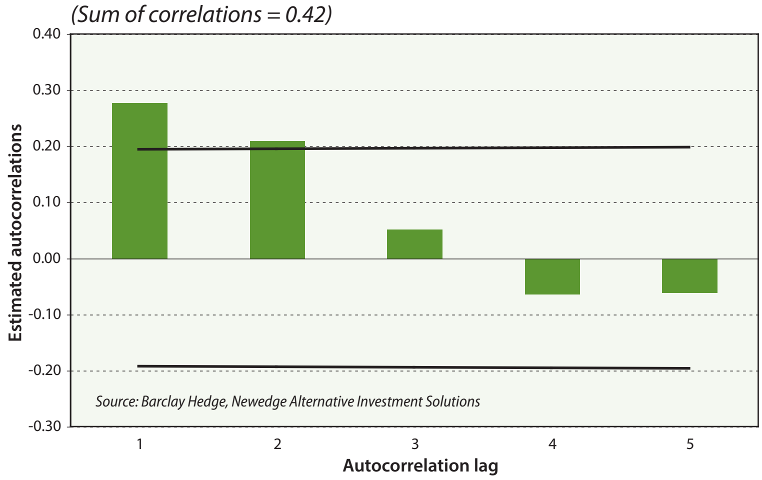


Exhibit 9
Autocorrelations for non-trend CTA #3
(Sum of correlations = 0.42)



volves our results for non-trend CTAs. In this case, the dots are at least scattered around the line in Exhibit 4, which is encouraging. But some of the dots are too far away from the line. The evidence for this is shown in Exhibit 6, which shows the non-trend CTAs' p-values for this exercise. What stands out here is that while the p-values seem to be stretched out along the diagonal from lower left to upper right, four of the 42 CTAs' had p-values that were 100% or only slightly less. And that's too many.

Estimates of autocorrelation

To be honest, autocorrelated returns were not on our original list of things that might explain what's going on in these two puzzles. More likely, we thought, the cause would be in variable volatility or something else that was not stationary. But none of what we tried produced any useful results. And so we turned to autocorrelation and seem to have found a plausible answer.

For example, consider the autocorrelation pattern for Trend follower #9 (TF9), whose p-value was next to lowest. We chose #9 partly because of its low p-value and partly because this CTA's maximum drawdown was furthest from its expected value in Exhibit 4. When we estimated the correlations between one month's returns and returns from one to five months earlier (say, TF9's June return with TF9's May, April, March, February, and January returns), the pattern of estimates we found was what you see in Exhibit 7. The first correlation was positive, but the rest were all negative. To provide some idea of statistical significance, we have shown 2-standard-deviation lines for the length of track record we used for this CTA. As you can see, the negative values were all significant or close to significant.

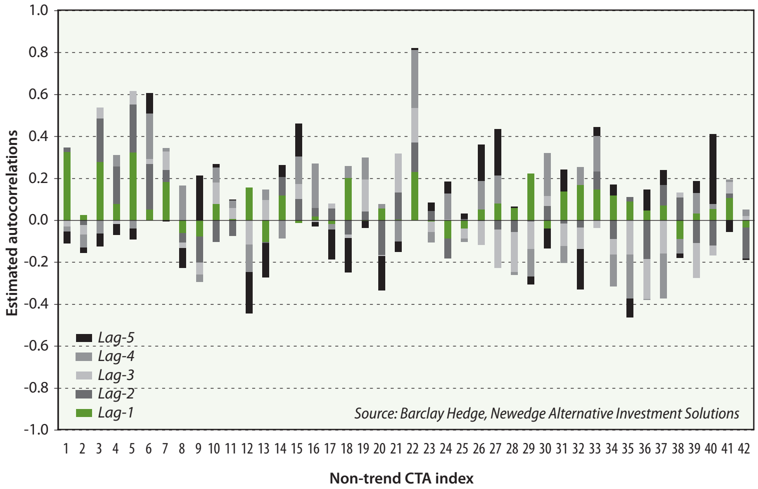
[Note: In what we report in this note, we limited our estimates to five lags, mainly because for most of the CTAs we examined, the size and significance of estimates beyond five lags tended to be small.]

What we found when we estimated autocorrelation structures for the other trend followers is shown in Exhibit 8. The vertical bars comprise shorter bars that represent the value of the autocorrelation value for each of the five lags. The graphic is useful in two ways. First, it shows that the patterns of positives and negatives could vary across CTAs. Second, and more importantly, it is easy to see that the net autocorrelation values – that is, the sum of the negatives and positives – were mainly negative for the trend following CTAs.

In contrast to Trend follower #9, consider what we found for Non-trend follower #3 (Non-TF3) in Exhibit 9. The p-value for this CTA was 1.00, which means that this CTA's realized maximum drawdown was far greater than our theoretical maximum drawdown distribution would have suggested was possible. And, when we estimate autocorrelations for Non-TF3, we came

Exhibit 10

Autocorrelations for non-trend components



up with the pattern shown in Exhibit 9. In this case, the first two correlations appear to be both positive and statistically significant.

What we found when we estimated autocorrelation structures for the 42 non-trend followers is shown in Exhibit 10. In this case, there is no systematic bias, at least in terms of sign. Some are positive. Some are negative. And so the fact that realized drawdowns for the non-trend set were scattered around the expected drawdown line in Exhibit 4 makes sense.

The scandal of the square root of time rule

At the heart of this note is the industry's failure to take autocorrelation into account when converting estimated daily, weekly, or monthly volatilities into the annualized volatilities that are used for risk assessment. In practice, the convention is to calculate, say, a daily price or return volatility (that is, the standard deviation of daily price changes or returns) and multiply the result by the square root of the number of business days in a year – usually something in the neighborhood of $256^{1/2}$. If one is using weekly price changes or returns, one would use the square root of 52. And with monthly returns, the multiplier would be the square root of 12.

This square root of time rule comes from the fact that the variance of the sum of a bunch of random, unrelated variables is equal to the sum of the variances of those same random, unrelated variables. And if we think that the variance of returns is the same from day to day, week to week, or month to month, this would look like

$$\sigma_n^2 = n\sigma_1^2$$

where σ_1^2 is the variance of daily, weekly, or monthly returns, σ_n^2 is the annualized return variance, and n is simply the number of days, weeks, or months in the year. If you take the square root of both sides to produce the standard deviations that we describe as return volatilities, the results is the familiar square root of n transformation

$$\sigma_n = \sqrt{n}\sigma_1$$

This transformation only works, though, as long as the returns are uncorrelated with one another from day to day, week to week, or month to month. But if they are not, the correct way to translate a single-period variance into an n-period variance is this:

$$\sigma_n^2 \approx n\sigma_1^2\left(1 + 2\sum_{i=1}^{k}\rho_i\right) \qquad \text{for } n > k$$

where ρ_i represents the correlation of one period's returns with returns from 1 through k periods ago, and n is quite a bit bigger than k.

Typically, we assume that these correlations are zero – that the returns from one period to the next are independent of one another – and blow off the term in parentheses. We use a squiggly equals sign to alert the reader to the fact that this relationship is an approximation that improves as the sample size (n) gets large relative to number of autocorrelation lags (k). In practice, each of the autocorrelation values is multiplied by a value equal to $(n\text{-}i)/n$, where i is the number of the lag (in our case, somewhere between 1 and 5). With 5 monthly lags, the effect of this weighting is to reduce the value of 2.0 to 1.75 if one has 24 months of data. If one has 60 months of data, the 2.0 would be reduced to 1.90. But with the autocorrelation values we have found in this round of work, the standard square root of time rule produces annualized volatility estimates that are either much too high or much too low. Consider what happens to annualized volatility calculations for different values of ρ_i. In Exhibit 11, we compare the

Exhibit 11

The wedge that autocorrelation drives between conventional and correct calculations of annualized volatility

Monthly volatility	Sum of autocorrelations	Autocorrelation adjustment to volatility	Annualized volatility without AC	Annualized volatility with AC
4.33%	-0.5	0.00	15%	0.00%
4.33%	-0.4	0.45	15%	6.71%
4.33%	-0.3	0.63	15%	9.49%
4.33%	-0.2	0.77	15%	11.62%
4.33%	-0.1	0.89	15%	13.42%
4.33%	0	1.00	15%	15.00%
4.33%	0.1	1.10	15%	16.43%
4.33%	0.2	1.18	15%	17.75%
4.33%	0.3	1.26	15%	18.97%
4.33%	0.4	1.34	15%	20.12%
4.33%	0.5	1.41	15%	21.21%

Source: Barclay Hedge, Newedge Alternative Investment Solutions

volatilities that one would get using an estimated monthly volatility or standard deviation of 4.33% (which is 15% divided by the square root of 12).

The size of the wedge that autocorrelation drives between the conventional and correct volatility calculations are shown in Exhibit 11 as the autocorrelation adjustment to volatility, which we have calculated by plugging the various sums of autocorrelations into the expression (1 + 2 x sum of autocorrelations) and taking the square root. The 0.00 adjustment ratio that goes with a correlation of -0.5 highlights the fact that the theoretical lower limit when autocorrelations is negative is -0.5. Notice that the only case for which the conventional and correct volatility calculations are the same is the one for which the sum of the autocorrelation values is zero. This is when the independence assumption holds, so this is when the standard square root of time rule works.

Exhibit 12

Maximum drawdown distributions for trend #9

(track record = 321 months, mean = 17.9%, volatility = 34.4% , observed max DD = -43.61%)

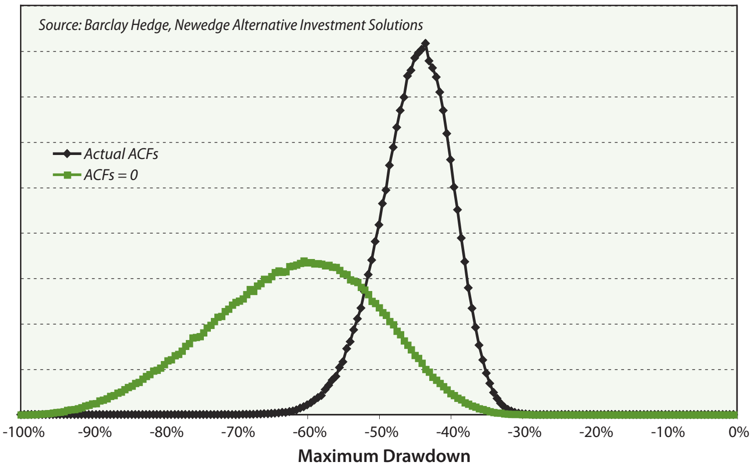
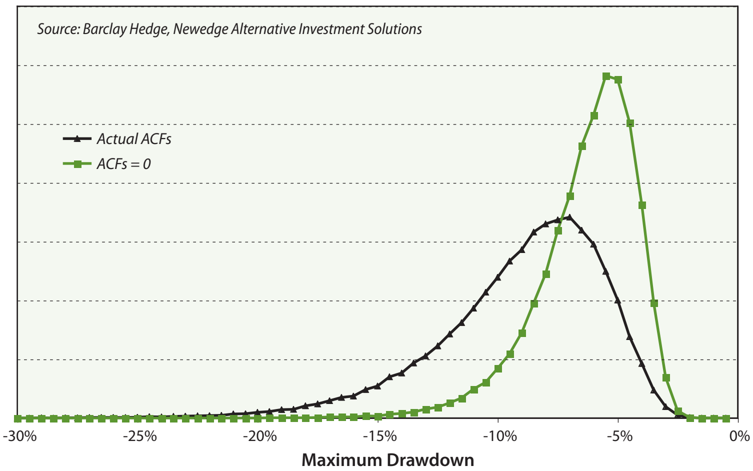


Exhibit 13

Maximum drawdown distributions for Non-trend CTA #3

(track record = 108 months, mean = 7.5%, volatility = 5.3% , observed max DD = -14.8%)



The table explains why the volatility calculation for TF9 was too high. In this case, the correlations summed to -0.42, which means that the correct value for TF9's volatility would be less than 0.45 of what the standard square root of time rule would yield. In other words, the standard approach produces a volatility that is more than 2 times too large.

On the other hand, the sum of the autocorrelation factors for Non-TF3 was +0.42. In this case, the adjustment would be a little over 1.34, which means that the correct volatility estimate for this CTA should be more than 34% larger than what is produced by the standard approach.

Second puzzle resolved

We are now ready to apply what we have learned about autocorrelation to our drawdown model. Exhibits 12 and 13 show the effect of including autocorrelation factors in our models.

In the first instance, we have drawn two maximum drawdown distributions for TF9 – one without allowing for autocorrelation and one in which autocorrelation is taken into account. Our original drawdown model would use length of track record (321 months), mean return (17.9%), and annualized volatility of returns (34.4%) as its inputs. And with these values, the maximum drawdown distribution is centered around 60%. When we incorporate negative autocorrelation, though, the drawdown distribution is shifted quite a bit to the right and is now centered around a value slightly larger than 40%. And with this distribution, the observed maximum drawdown of -43.6% appears to be more comfortably in the middle.

When we work with Non-TF3, we find that the

Exhibit 14

Expected and observed maximum drawdowns for components of the Newedge CTA Index with allowance for aucorrelations

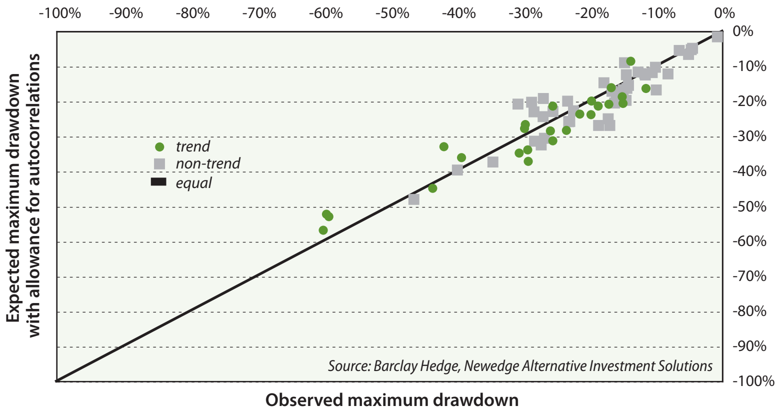


Exhibit 15

Correlation between Newedge Trend Subindex and mini trend index

(January 2000 through July 2012, rho = 94.7%)

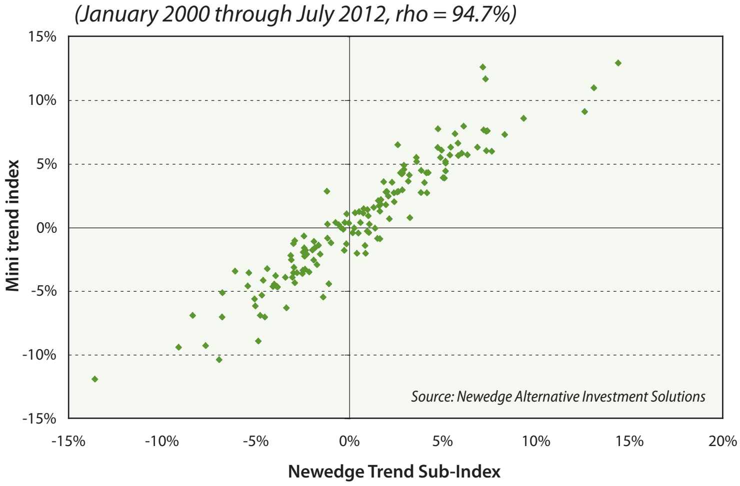
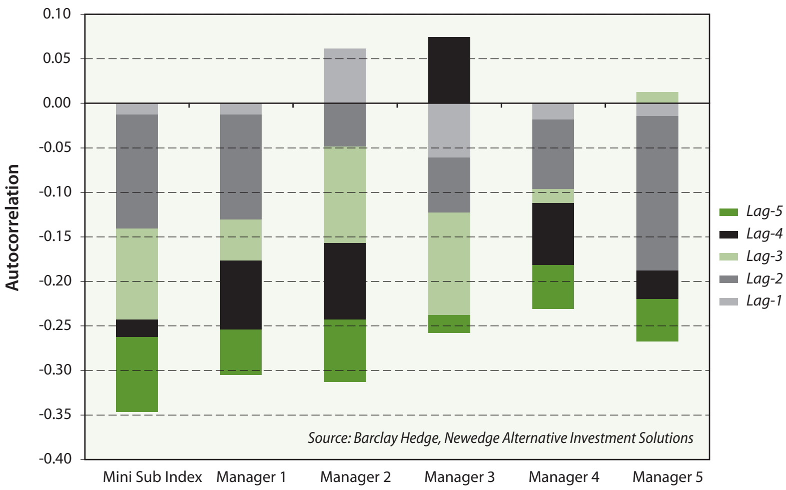


Exhibit 16

Autocorrelations for the mini trend index and its components



presence of autocorrelation shifts the maximum drawdown distribution noticeably to the left. Using inputs of track record (108 months), mean return (7.5%), and annualized return volatility (5.3%), we find that the distribution is center somewhere close to 5%. When we allow for positive autocorrelation, however, the distribution shifts out and is centered over something closer to 8%. And this manager's maximum drawdown of -14.8%, while it is toward the upper end of the distribution, is more probable than it would have been with the original distribution.

We also find that the empirical puzzle presented in Exhibit 4 is mainly corrected when we use the new drawdown model to reckon expected maximum drawdowns. In Exhibit 14, we have plotted expected versus realized maximum drawdowns for the original set of CTAs, but this time we have used their respective autocorrelation estimates when calculating the expected maximum drawdown for each.

This approach produces two good outcomes. First, the trend followers' values are now scattered above and below the line rather than almost entirely under the line. Second, the non-trend followers' values are now scattered closer to the line. In both cases, the resulting differences between experience and expectation exhibit more reasonable variability.

Extending the work to a broader data set

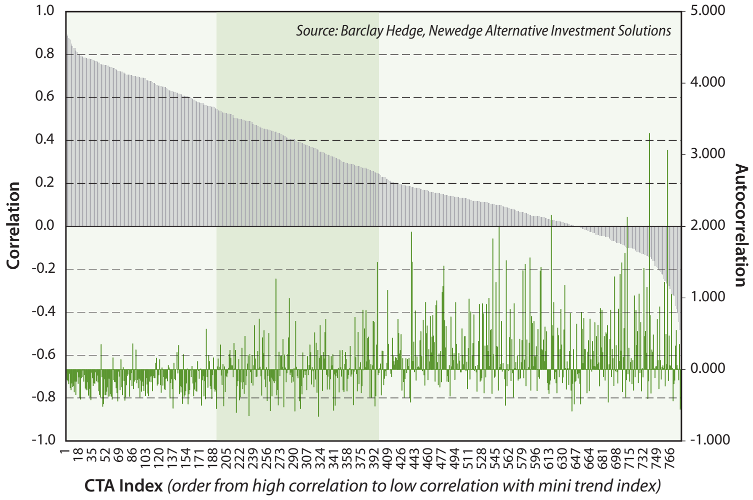
To address the question of whether what we found in the first round of work was an accident of the data, we tackled the problem of working with a much more comprehensive set of data for a total of 783 CTAs.

Working with this set posed a special challenge because we could not know each of the 783 CTAs as intimately as we knew the 67 larger, well established CTAs who had been part of the Newedge CTA Index. At the same time, it allowed us a chance to approach the problem of classification in a slightly different way. And, as a result, we uncovered more useful insights into the relationship between trend following and autocorrelated returns.

The first step in working with this data set was to construct what we call a "mini trend index" that could serve as a benchmark for trend following behavior. To do this, we identified five CTAs who have always been known as trend followers and whose track records were available for an extended history. As it was, we were able to start this mini trend index in February 1988 and run it through July 2012, which is the last month used in this work. As a reasonableness check on whether this mini trend index could serve as a legitimate proxy for trend followers, we plotted monthly returns for the mini trend index against monthly re-

Exhibit 17

Sum of first five autocorrelations for CTAs



turns for our Newedge Trend Sub-index and found a correlation of 94.7%. The scatter is shown in Exhibit 15 and shows monthly returns from January 2000, which is the first month for which the trend subindex data were available. And, out of curiosity, we calculated the autocorrelation factors for the five CTAs used in the mini trend index and came up with the results shown in Exhibit 16. For the index itself and for two of the component CTAs, all of the autocorrelation factors were negative. For three of the CTAs, one of the five lagged correlations was positive. But on balance, the autocorrelations for all five were negative.

We then calculated the correlation of each of the 783 CTAs' monthly returns with those on the mini trend index and ordered them from highest to lowest. The next step was to use each CTA's own track record to calculate autocorrelation factors. The results of these two exercises are overlaid on each other in Exhibit 17. The sum of the first five autocorrelation factors is measured on the left vertical axis. The correlation of each CTA with the mini trend index is measured on the right vertical axis.

The results are really quite astonishing. The CTAs with the highest correlation to the mini trend index are on the left and consistently exhibit negative autocorrelations. This seems to be true for the first 200 or so CTAs on the left. Then, as correlation with the mini trend index falls, the sum of a CTA's autocorrelation factors tends to rise. A visual scan of the results suggests that the next 200 CTAs are distributed more or less evenly around the zero autocorrelation line. Then, when correlation with the mini trend index falls to 0.25 and below, the sums of the remaining CTA's autocorrelation factors are mainly positive.

A quick comment about the sizes of the autocorrelation values is probably in order. The negative numbers look smaller than the positive numbers, but in fact there is a kind of asymmetry here. First, as noted above, it is important to know that the theoretical lower bound for autocorrelations is -0.5. Anything smaller than this would imply a negative variance, which might be interesting but is not possible. So an average of roughly -0.25 for the CTAs who would be grouped as trend followers is roughly halfway between zero and the theoretical limit. And if we apply this number using

$$\sigma_n^2 = n\sigma_1^2\left(1 + 2\sum_{i=1}^{k}\rho_i\right)$$

we would find that the variance (not the volatility) would be just half of what the normal expansion for single-period to multi-period would produce.

On the positive side, the theoretical upper limit of the sum for five autocorrelation factors would be +2.5 – which would represent a correlation of 0.5 for each of the five factors. As it is, there are some stray large values, some of which exceed the theoretical upper limit. These can be the result of sampling error or variable volatility in the underlying series, but they cannot be used in any applied work or simulations. Generally, however, for those CTAs with correlations to the mini trend index below 0.25, the average value was in the neighborhood of +0.50. And if we apply this number, we find that the variance would be double that we would get without considering autocorrelations. So at least in variance terms, values of half and double are mirror images of each other.

Autocorrelation and the Newedge Trend Indicator

The Newedge Trend Indicator is not a CTA, but it does represent a trend following model that correlates well with the returns of well recognized trend followers. So, as a final piece of evidence, we calculated the autocorrelation factors for this model's returns. The results of this exercise are shown in Exhibit 18.

Exhibit 18

Autocorrelations for Newedge Trend Indicator
(January 2000 through July 2012)

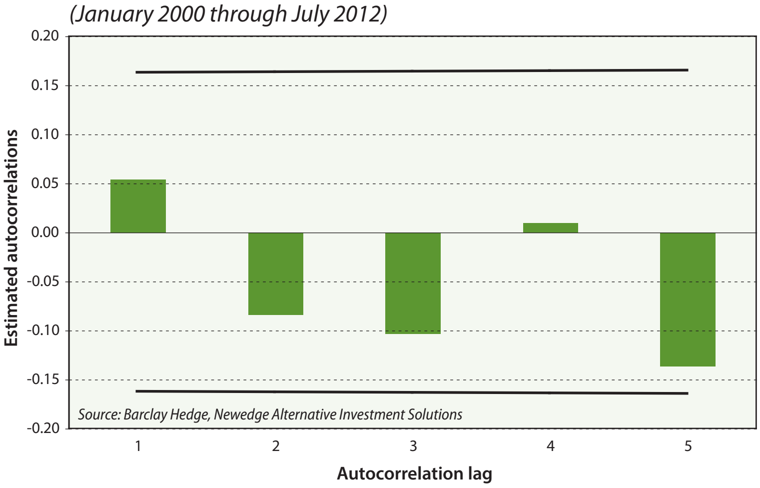


Exhibit 19

60-month rolling autocorrelations for mini trend index components

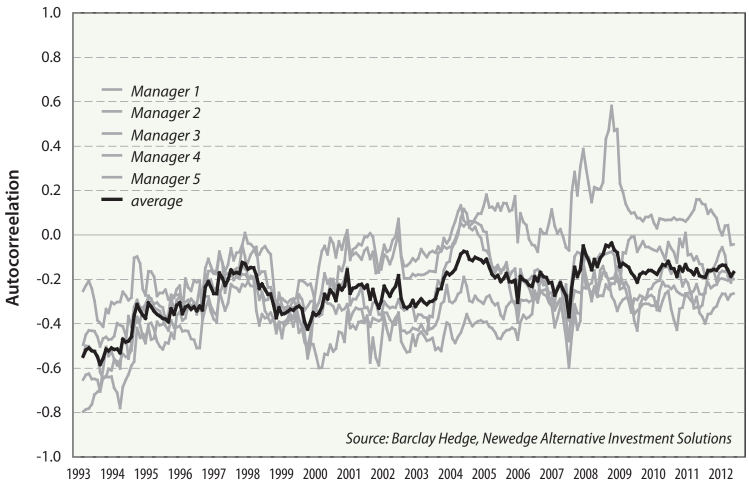
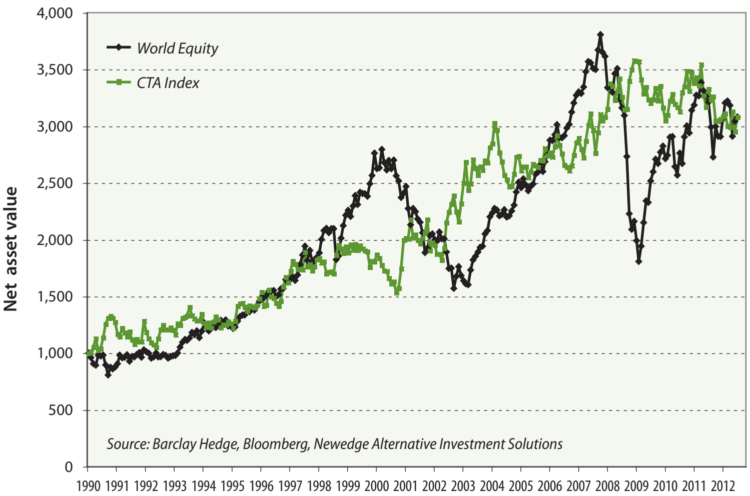


Exhibit 20

Net asset values for two return series
(annualized mean = 5%, annualized vol = 15%)



Although none of the factors is significantly negative, the pattern of mainly negative values suggests that the force is at work here as well.

The question of persistence

As it is, we have found evidence of significant autocorrelation in all of the data we have been able to assemble. There remains, however, the question of whether it is persistent. To address this question, the best data set we have comprises the returns of the five trend followers that we used to construct the mini trend index. The advantage of this set is it length and consistency. We have uninterrupted return series for all five going back to 1988, and so we don't have to worry about changes in the composition of our data.

For these CTAs, we used 60-month rolling periods to calculate the average value of each autocorrelation factor from lag 1 to lag 5, and from these we also calculated the sums. These six time series are charted in Exhibit 19. One notices at least two things in these histories. First, the estimated values can vary quite a lot over time, and, in some instances can be positive for an individual manager. Second, and most important, the sum of the five individual sums has never been positive and seems to have been roughly stable in the area of -0.20, plus or minus, from the late 90s on.

The choice of a 60-month rolling estimation period was influenced in part by the time it takes to detect autocorrelation with any kind of statistical reliability. The standard deviation of a single correlation estimate when the true correlation is zero is roughly one over the square root of the number of observations, or $n^{-1/2}$. With 60 months, the standard deviation would be about 0.13 [= $60^{-1/2}$], which means that five years or 60 months is about the amount of time needed to detect an overall correlation of 0.25.

Why investors should care about autocorrelation

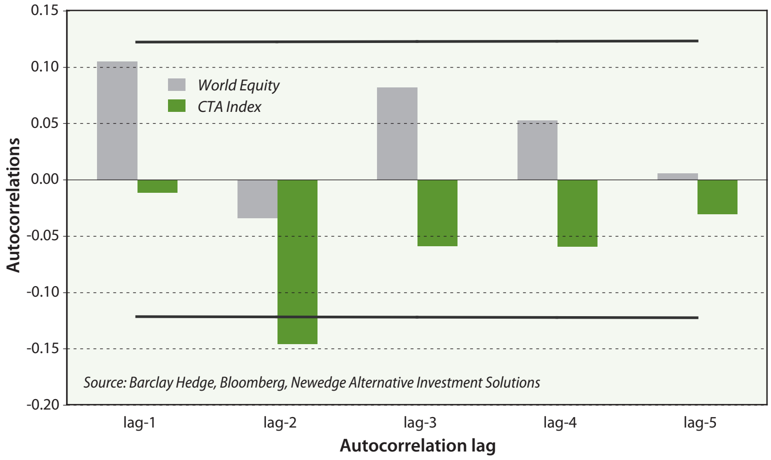
Perhaps the single most important reason to pay attention to autocorrelated returns is that if we do, we can get better ideas of the riskiness of various assets in our portfolios. If we return to the normalized net asset value series for world equities and CTAs shown in upper panel of Exhibit 1, which we have reproduced here as Exhibit 20, we can reconsider the way we compare the two series.

The eye tells you that these two series represent entirely different kinds of risk. And if we estimate the autocorrelation structure for the two, what we find is shown in Exhibit 21. The autocorrelation values for world equities are generally positive and sum to +0.21. In contrast, the autocorrelation values for CTAs are generally negative and sum to -0.31.

Exhibit 21

Autocorrelations for world equities and the CTA index

(Autocorrelation sum = +0.21 for world equities and -0.31 for the CTA index)



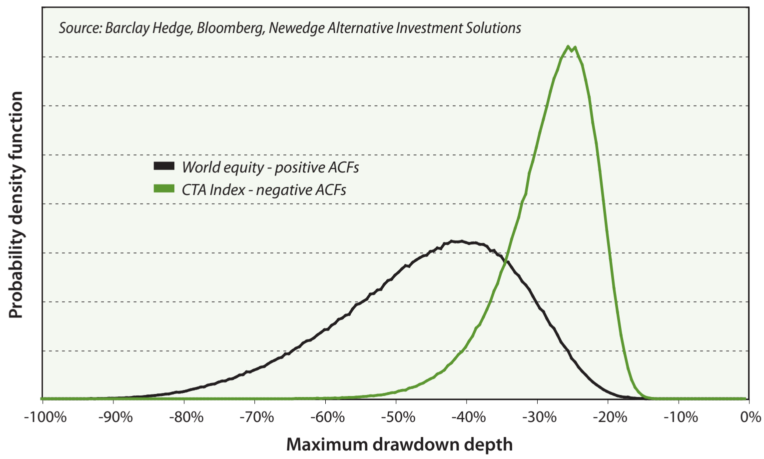
And so, even though the standard square root of time rule produces annualized volatilities of 15% for both series when applied to monthly volatilities, we know that the correct volatilities are higher for equities and lower for CTAs. Knowing this, our expectations for drawdowns are much different for the two. As shown in Exhibit 22, it is now perfectly plausible that equities could produce 50% drawdowns as they did in the past decade.

It also makes sense that we should revise the way we calculate risk-adjusted returns. As shown in Exhibit 23, the ratios of return to risk would be 0.33 for both equities and CTAs if we ignore autocorrelation. But we see that the correct volatility estimate for equities is really just over 18%, while the correct volatility estimate for CTAs is just under 10%. As a result, if we calculate the return/risk ratios taking this into account, we find that the ratios are far from the same. For equities, the ratio is now 0.28, while for CTAs, the ratio is 0.53 – almost double that of equities.

Another way to think of the riskiness of the two assets is reflected in Exhibit 24, which shows the results of running the following experiment, which would be relevant for a defined benefit plan. In each case, we started with $1,000 and then withdrew $50 a year (or really $4.17 a month) to pay a fixed cash obligation. The $50 corresponds to the 5% return assumed for each asset. We then simulated the distribution of outcomes for these two assets over a horizon of 270 months, which corresponds to the histories we have used in this note.

Exhibit 22

Distributions of maximum drawdown depths for return series with non-zero autocorrelations



The two distributions are instructive, in part because they drive home the point that if insolvency or bankruptcy is a possibility, then standard deviations alone are not an adequate measure of risk. In this exercise, insolvency occurs whenever the value of the asset drops below $4.17 and the fund cannot meet its obligation.

We also see that an investor's evaluation of risk requires some estimate of the costs of insolvency. In the equities example, the probability of insolvency was 21.7% [= 100% - 78.3%, which is the probability of survival]. For CTAs, the probability of insolvency was 4.0% [= 100% - 96.0%]. On the other hand, the expected value of the asset if it is still solvent at the end of the period was $2,519 for equities but only $1,224 for CTAs. The difference between these two expectations may be part of the reason that so many defined benefit public employ pension plans are invested so heavily in equities.

Exhibit 23

Effects of autocorrelation on performance measures

		Annualized mean	Annualized volatility	Return/risk ratio
Without autocorrelation corrections	Series A	5%	15.0%	0.33
	Series B	5%	15.0%	0.33
With autocorrelation corrections	Series A	5%	17.9%	0.28
	Series B	5%	9.4%	0.53

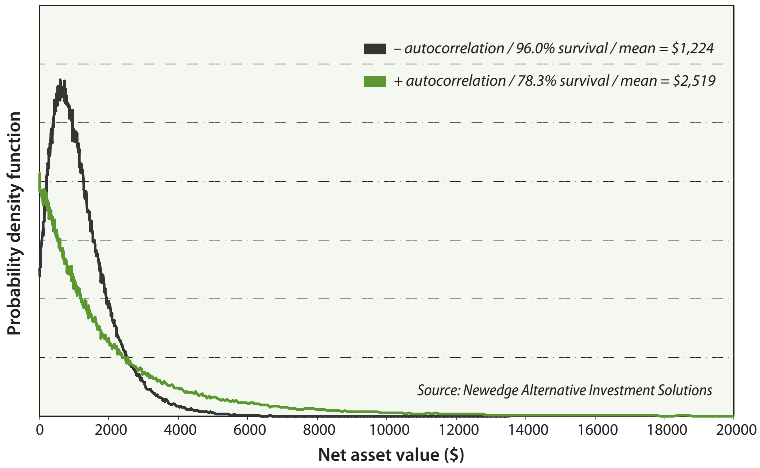
Source: Barclay Hedge, Bloomberg, Newedge Alternative Investment Solutions

Using autocorrelation as a diagnostic

For those investors who are interested in having the best available measures of risk at their disposal, the possibility of using autocorrelation estimates to detect biases in standard volatility measures is good news. At least it is fairly good news. The challenge is that detecting autocorrelation takes time, and the time it takes may be longer than many risk evaluation horizons allow.

On the other hand, the alternative is in some sense worse. One need not, in principle, use autocorrelation to get an unbiased estimate of an asset's riskiness. One could, instead, calculate independent,

Exhibit 24
Probability distributions for ending net asset values
(5% mean, 15% vol, 270 months)



non-overlapping n-period returns (where n is long enough to include all the correlation lags). But the time it would take would be much, much longer than is required to work with correlations. To see why, consider a simple 1-period lag. In such a case, one should be calculating 2-month returns. If they are to be non-overlapping, then one would get 6 observations per year instead of 12. But for them to be independent, one would have to skip a month between each 2-month return period. If you do this, you are down to 4 observations a year. And in our case, where the lags seem to stretch out to five months, one could really only get one independent observation a year. In which case, it would take a few decades to learn what would be available in only a few years if one were to use autocorrelations to augment the standard measures of volatility.

Where do we go from here?

We are persuaded at this point that the presence of autocorrelated returns helps to resolve the two empirical puzzles that prompted this work in the first place. At least one natural extension of this work, then, will be to amend our manager evaluation reports to reflect the new drawdown model and to include some information about autocorrelation in each manager's returns.

Timing investments But there are a number of questions that will require some attention. For one, we examined the question of whether one could time one's investments in CTAs using past returns, found generally that the answer was no, and published our findings in *Every drought ends in a rainstorm*. The main conclusions were that CTAs' returns appeared to be uncorrelated through time. Returns conditioned on past returns appear to be the same as unconditional returns. And the numbers of runs of gains and losses appeared to be thoroughly consistent with randomness. The data set was small, though, and we may arrive at different findings with the broader data sets we have used in this work.

Drawdown control If returns are autocorrelated, it is possible to improve one's risk-adjusted returns using what is loosely called drawdown control. For return series with positive autocorrelation, a rule that reduces position sizes when losing money and increasing position sizes when making money will improve risk-adjusted returns. If returns are negatively autocorrelated, the opposite should be true, in which case drawdown control would increase position sizes when losing money and would decrease position sizes when making money.

Where does autocorrelation come from? And why is it negative for trend followers? These are thorny questions but important to explore. In early conversations about this work, one of the questions raised was whether autocorrelation was evidence of inefficiency – of money being left on the table, or money lost unnecessarily. For that matter, part of the reasoning we used in *Every drought ends in a rainstorm* to explain the absence of conditionality in returns was that active trading, if well done, would wring returns from price series that exhibited positive or negative momentum and produce return series that were free from autocorrelation.

Postscript on statistical influences on drawdowns

In *Understanding drawdowns*, we found that the three most influential variables were length of track record, return volatility and mean return. Skewness, which indicates whether upside returns are larger or smaller than downside returns, didn't seem to have much effect. Neither did excess kurtosis, which measures the likelihood of unusually large positive or negative returns.

The reason is most likely the power of the central limit theorem, which states that the sum of enough independent random variables will be normally distributed, no matter how each of the individual variables is distributed. So one can add up the oddest, most peculiar, random variables, and the central limit theorem pounds them all into a fairly uniform, normally distributed paste. In fact, skewness and kurtosis probably do matter for CTAs with short track records, but only because the central limit theorem has not had time to work its erosive magic.

In contrast, autocorrelation does not disappear as the track record lengthens. If it's there, it's there, and becomes more noticeable with the passing of time, not less. The wedge it drives between estimates of volatility based on returns for single periods that are shorter than the length of the autocorrelated lags and of the asset's multi-period volatility simply becomes clearer and more pronounced.

Acknowledgements

We want to thank Mark Carhart of Kepos Capital for his contribution to this piece of work. At our September 2011 manager/investor forum in San Francisco, he presented some work on drawdowns that explored the possibility of using drawdown control to improve risk adjusted returns if, as is the case for certain classes of hedge funds, returns are positively autocorrelated. He also encouraged the use of drawdown per unit of volatility when comparing one manager with another. It was the latter suggestion that produced the empirical puzzle that you find in Exhibit 4. And it was his thinking about autocorrelated returns that led us to think of autocorrelation as the solution to the puzzle. In addition, he has met with us regularly over the past year to discuss the ways in which we might use drawdowns as a diagnostic tool when evaluating expectations about a manager's future and about the appropriateness of drawdown control as a risk management tool.

Any questions about the derivations behind the transformation shown on page 6, the theoretical limits on positive and negative autocorrelations, and the calculation of significance bands for autocorrelation estimates should be directed to Lianyan Liu.

Appendix: Normalizing world equity and CTA returns so that they have equal means and volatilities

The actual net asset value series for World Equity and CTAs are those shown in Exhibit 25. Over this period, CTA were the superior asset if compared solely on returns (higher than those on equities) and volatility (lower than that of equities). To isolate the effect of autocorrelation on drawdown behavior, however, we created two new series that have identical mean returns and return volatilities but that preserve the pattern or sequence of returns in each. That is, whatever autocorrelation one would find in the original series, one would also find in the modified series.

To do this requires only two steps. The first is to normalize the original, actual series so it has mean = 0 and volatility = 1 as follows:

$$R_{2t} = (R_{1t} - \mu_1)/\sigma_1$$

where R_{1t} is the original return for month t, μ_1 is the average return for series 1, and σ_1 is its volatility or standard deviation of returns. The resulting distribution of R_2 will now have a mean of zero and a standard deviation of 1.0, but the sequence of returns will preserve any aucorrelation that one would find in the original series. The second step simply requires that the second series be converted to a third series as follows:

$$R_{3t} = \mu_3 + \sigma_3 R_{2t}$$

Where μ and σ can be any values you choose. In this note, we arbitrarily chose a mean of 5%, which roughly splits the difference between the observed mean returns for CTAs and World Equities, and a volatility of 15%, which looks more like equity volatility. These are the series that we show in Exhibits 1 and 20. In the end, we have two series that would have identical means and volatilities (and, as a result, identical risk-adjusted returns), so that any difference in drawdown behavior can be attributed only to the pattern of returns.

Exhibit 25

Net asset values for two return series