The lights are on

Christopher C. Finger
chris.finger@riskmetrics.com

October 2006

A month after we wrote in this space that essentially, nothing exciting had happened in the markets this year, we had Amaranth and its $6 billion of losses, much due to trading in natural gas futures. Though the media attention has now waned, the episode did bring its share of hand wringing. There are numerous summaries of the public version of the events. It is thus not our goal here to summarize all that is known about Amaranth, but rather to clarify one point in the aftermath.

Predictably, highly publicized losses have led to reexaminations of the oversight that existed (or did not exist) on the fund's activities, and predictably, some of these reexaminations have led to the "throw out the models" conclusion. Among the more provocative statements is the following:[1]

> . . . using them (VaR models) to understand the financial risk of a spread trade in a physical commodity is like driving really fast at night with your headlights turned off.

Unfortunately, the author of this statement presents neither data nor a solution.

Moreover, in the abstract of Till (2006), presumably to grab attention, is the statement that the Amaranth losses were a nine-standard deviation event. Unfortunately, it is this provocative statement, rather than the lessons in the body of this otherwise thorough and timely analysis, that has grabbed the most media attention.

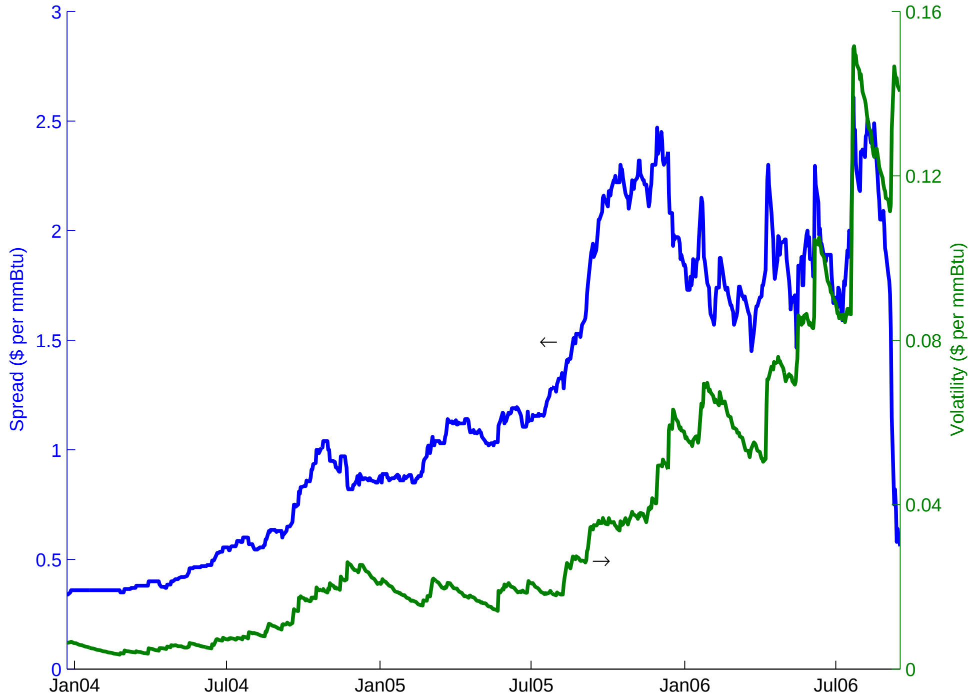
While we have preached transparency and "knowing what you don't know" repeatedly, this is a time to defend the models. There have been market crises in the past (for instance, currency pegs breaking, some equity crashes, or the spike in the swap spread in 1998) that produced even more astounding events. With these, there is often nothing to conclude other than that there was no information in the models that was at all useful in warning of the impending danger. Sometimes the lesson is a bit more helpful, such as in the structured credit stresses of last year,[2] and we learn that our risk models simply need to incorporate more factors.

Here, as we will see, the models did what they were designed to do, and no one should have been as surprised as the driver who runs into a tree they did not see. We should acknowledge this, and not let hysterical comments distract us from lessons we can learn about modeling choices and liquidity monitoring.

[1] From *Credit Risk - Amaranth Aftermath* at www.RiskCenter.com

[2] See Finger (2005).

Figure 1: Natural Gas NYMEX futures contract. Spread between prices for delivery in March 2007 and April 2007, and spread volatility based on EWMA



Natural gas and volatility

Before delving into specific trades, we recall the first lesson in Till (2006):

> …an examination of the fund's monthly sector-level profit-and-loss would have been sufficient to raise a red flag.

Amaranth's energy strategies had shown recent monthly gains and losses of 12%; one should therefore have not been surprised by moves roughly twice as large as these (two-standard deviation events). So even the simplest model was somewhat informative.

As has been well reported,[3] some of Amaranth's largest positions were in calendar spread trades on natural gas. In one such trade, they held a long position in gas for delivery in March and a short position for delivery in April. The position could be for this year or future years; there are futures contracts traded on NYMEX for delivery as 2011. The position makes money if the spread between March and April gas increases and lose money otherwise.

In Figure 1, we plot the evolution since 2004 of the March-April spread on NYMEX futures contracts for delivery in 2007; we also plot the volatility forecast for the spread, computed using an exponentially weighted moving average (EWMA) with decay factor of 0.97. We see that the spread increased in late 2005, came back down, and spiked upwards in late July of 2006. Over the week of September 11, and particularly on the day of September 15, the spread

[3] See Till (2006) and references therein.

Table 1: Risk forecasting results for March-April 2007 Natural Gas spread

	Residual	p value	Loss/VaR	Loss/ES
September 15				
EWMA, Gaussian	-3.54	0.02%	1.52	1.33
EWMA, *t*	-3.54	0.30%	1.36	1.03
Equal, Gaussian	-4.63	0.00%	1.99	1.74
Equal, *t*	-4.63	0.09%	1.78	1.34
September 11–15				
EWMA, Gaussian	-2.93	0.17%	1.26	1.10
EWMA, *t*	-2.93	0.65%	1.12	0.85
Equal, Gaussian	-8.75	0.00%	3.76	3.28
Equal, *t*	-8.75	0.00%	3.36	2.54

plummeted, causing disastrous losses on the long spread positions.

Examining the volatility of the spread, we see a very consistent level, even through the sizeable run-up and subsequent decrease in late 2005. From early 2006, we see a protracted increase in volatility: by August, the spread's volatility was roughly three times greater than it was in January.

Forecasting risk

For most risk models, there are two basic decisions: the technique to forecast volatility, and the distribution of the residuals (that is, the returns normalized by the volatility forecast). Intuitively, the first decision tells us how large an event is in units of standard deviations, and the second tells us how rare (or likely) such an event is. Under the simple assumption that the expected return is zero, the VaR forecast is the product of the volatility forecast and the relevant quantile of the residual distribution.

We look at two standard methods to forecast volatility: the EWMA mentioned above, and the equally-weighted standard deviation over the previous year. For longer risk horizons than one day, we scale our daily volatility forecast by the square root of the horizon length.[4] We consider also two choices for the residual distribution: the Gaussian, and Student's t distribution with five degrees of freedom. The t distribution is a natural choice for any security where we expect to observe fat tails. Five degrees of freedom is by no means an extreme choice, but rather one that fits well across a broad range of asset classes.[5]

We apply these simple risk modeling choices to the Natural Gas spread, and examine the performance of the risk forecasts for the day of September 15 and the week of September 11–15. We calculate the fol-

[4] Though common practice, this scaling is best considered a convention and a first order approximation. See Zumbach (2006) for a discussion of alternate approaches.

[5] See Zumbach (2006).

lowing statistics:

- Residual—the actual loss divided by the volatility we would have forecast using information available at the time,
- *p* value—the probability, according to our model, of observing a return as bad or worse than what actually occurred,
- Ratio of the actual loss to the VaR (at 99% confidence) forecast, and
- Ratio of the actual loss to the Expected Shortfall (also at 99% confidence) forecast.[6]

The results are presented in Table 1. Using the EWMA forecasts, we see that the loss of September 15 was only about a 3.5-, and the loss of September 11–15 a three-standard deviation event. These are significant numbers, but a far cry from a headline-grabbing nine-standard deviation event. Under the Gaussian distribution, these events should still be rare: about one-in-5000 for September 15 and one-in-500 for September 11–15. Under the *t* distribution, the *p* values suggest that September 15 was about a one-in-300, and September 11-15 a one-in-150 event. Examining the results from a different angle, we see that the loss of September 15 was between 36% and 52% greater than our VaR forecast, and the loss of September 11–15 between 12% and 26% greater than VaR. In fact, in both cases, the losses were no greater than the ES forecasts.

So the losses were large, but the statistics should lead us to adjectives like *unlikely*, but not to *rare*, *aberrant*, *unthinkable*, or *exceptional*. Of course, had we chosen an ill-advised (that is, equally-weighted) volatility forecast, particularly for September 11–15, we would have been left with close to a nine-standard deviation event, and all of these strong adjectives.

From the modeler's point of view, it would be nice if the story ended here, but we should acknowledge that the 2007 (that is, current year) contracts are the best behaved of the natural gas contracts that trade. Contracts for delivery further into the future trade less liquidly and from a statistical point of view, are much more difficult to forecast.

We apply the same modeling as before to the March-April spread, but for 2008 delivery. The results are presented in Table 2. Here, the story is worse. Even with the EWMA model, the events in question are seven- to nine-standard deviation events, translating under the *t* distribution to a one-in-8000 to one-in-30,000 likelihood. Even so, for September 15, the loss was less than four times the VaR forecast, and less than three times the Expected Shortfall forecast; and for September 11–15, the loss was less than three times the VaR and about twice the ES. Moreover, the risk forecasts were almost twice in September what they had been in January: the market heated up, and the forecasts reflected that.

For one more perspective, consider that under the 1998 Market Risk Amendment to the Basel Accord, banks may calculate capital as the 99% VaR times a capital multiplier. The capital multiplier, set to protect against extreme events, is usually three or five.

For the week of September 11, the spread move was less than three times the VaR forecast, meaning that the smaller multiplier would have been sufficient to ensure solvency. For the two-week[7] pe-

[6]The Expected Shortfall (ES) is the expected loss, given that the loss is greater than VaR.
[7]The standard risk horizon for setting trading book capital

Table 2: Risk forecasting results for March-April 2008 Natural Gas spread

	Residual	*p* value	Loss/VaR	Loss/ES
September 15				
EWMA, Gaussian	-9.44	0.000%	4.03	3.54
EWMA, *t*	-9.44	0.003%	3.62	2.74
Equal, Gaussian	-11.15	0.000%	4.76	4.18
Equal, *t*	-11.15	0.001%	4.28	3.23
September 11–15				
EWMA, Gaussian	-7.22	0.000%	3.08	2.71
EWMA, *t*	-7.22	0.012%	2.77	2.09
Equal, Gaussian	-18.00	0.000%	7.69	6.75
Equal, *t*	-18.00	0.000%	6.91	5.22

riod beginning September 11, the spread move was 3.6 times the VaR forecast, meaning that at least the larger multiplier would have guaranteed solvency. So while the moves were large, and did surprise our simple model, they were well within the realm of the large moves that policies built on models are meant to address.

The L word

We conclude, then, that the standard risk models performed adequately in this case, giving at least some warning of the scale of possible losses. But a second question is whether the losses addressed by the standard models were all of the story.

Here, we must admit that the answer is no. Standard risk models (and our examples specifically) forecast based on information about the closing prices of the securities in question. There is no way for such a model to address the event where a market participant, holding a very large position, sells that position under pressure at a significant discount, and then the market rebounds and closes at a reasonable level. Yet by all accounts, this is what happened in the Amaranth case, and as Till (2006) points out, likely in the case of MotherRock (another energy-related hedge fund) as well.

One thing the models are missing, then, is intraday information. One way to address this is to model based not on closing prices, but on volume-weighted average prices. Here, we only observe one price point per day, but that price is driven by where most of the trading volume lay, rather than by what turned out to be the last trade. A second direction is to sample prices much more frequently. This can improve daily or weekly forecasts, but the benefits fall off for longer horizons; and sampling more data clearly brings a cost in model complexity.

Under both of these approaches, however, we remain rooted to the assumption that an investor is a victim of the market volatility, rather than a protagonist in it. As long as the positions we consider are small relative to the market's volume, we can rest assured that our decision of whether to liquidate our positions or merely value them based on market quotes will have no impact on the actual prices. With positions that are large relative to the average trading volume, we must consider the effect of our activity on the very prices from which we model.

Ultimately, we would like a risk model that encompasses forecasts of general market moves and potential liquidity shocks that could result from our specific positions. Such a risk model has thus far proved elusive, but its absence should not deter us from at least some straightforward monitoring of liquidity. For exchange-traded products (equities, bonds, futures contracts, and some derivatives), we may observe the average volume of trading, and track the size of our positions relative to this volume. This at least identifies those positions that pose the greatest potential liquidity risk—to which we should apply greater scrutiny, stress tests, or curtailment—even if we are unsure of how to quantify that risk, or connect that risk with moves in the general market.

While the comments above deal with endogenous liquidity risk (that is, risk deriving from our own specific holdings), there is also a lesson here about exogenous liquidity risk—the risk deriving from the overall liquidity, or lack thereof, in a security. The overall liquidity of markets should influence our choices of which price time series to model. As we observed, risk forecasts for the 2007 spread performed reasonably well, while those for the 2008 spread were less than ideal. This is not surprising, given that the market for the 2008 contracts was thinner, making for greater possibilities of sudden price moves. The lack of general liquidity in the 2008 (not to mention the 2009 or 2010) contracts begs the question of whether it was appropriate at all to forecast based on the prices of these contracts. It would have been defensible to proxy the risk of all the longer contracts with their 2006 or 2007 counterparts. This is a tradeoff we have written about numerous times: we gain forecasts based on well-behaved time series, but lose the direct connection from the time series to our positions.

In this specific case, with the benefit of hindsight, the choice would have worked out well: though the 2007 and 2008 spreads were only moderately correlated through most of 2006 (57% through the end of August), they exhibited nearly identical moves through the stresses of September. Using the proxy, the moves in the 2008 spread would have only exceeded the VaR forecast by only 60% and 12%, for September 15 and September 11–15, respectively.

Final thoughts

Before making any concluding remarks, it is important to appreciate the context of the market moves we have examined. The events of September led to the greatest losses ever by a single hedge fund, close to twice the money lost by Long Term Capital Management. If we assume that there have been on the order of 10,000 hedge funds, with an average track record of five years, then there have been roughly 2.5 million weekly hedge fund returns, and we are looking at the worst. These were rare events, and we should be worried (particularly if we are invested in other hedge funds) if our risk models shrug them off and characterize them as one-standard deviation occurrences.

But these were not events that occurred without warning signals, and even a standard model would have shed light on the scale of the potential losses.

Granted, there were also losses that standard models did not address, but there are better lessons to take away than throwing the models out the window: whether investigating volume-weighted price data, reassessing our use of proxy time series, or monitoring developing liquidity problems.

Maybe we were driving with the low-beams only, but anyone using a reasonable risk model at least had the headlights on. There are certainly bad model choices, as the performance of the equally-weighted estimation demonstrates here, but there are also good model choices, even fairly simple ones. The problem, alas, is sometimes not the car, but the driver.

Further reading

- Finger, C. (2005). Eating our own cooking, Research Monthly, June.
- Till, H. (2006). EDHEC Comments on the Amaranth Case: Early Lessons from the Debacle. EDHEC Risk and Asset Management Research Centre.
- Zumbach, G. (2006). A gentle introduction to the RM 2006 methodology. Working paper. RiskMetrics Group.