SOCIETY OF ACTUARIES

Article from:

Risks and Rewards Newsletter

March 1998 – Issue No. 30

Measuring Financial Risk
continued from page 17

in the portfolio return distribution also represent a problem for the parametric approach. Historical simulation (my personal favorite) is free from distributional assumptions, but requires the portfolio be revalued once for every day in the historical sample period. Because the histogram from which the VAR is estimated is calculated using actual historical market price changes, the range of portfolio value changes possible is limited. Monte Carlo VAR is not limited by price changes observed in the sample period, because revaluations are based on sampling from an estimated distribution of price changes. Monte Carlo usually involves many more repricings of the portfolio than historical simulation and is therefore the most expensive and time-consuming approach.

Rule or Tool?

It seems that VAR is being used for just about every need: risk reporting, risk limits, regulatory capital, internal capital allocation, and performance measurement. Yet, VAR is not the answer for all risk management challenges. No theory exists to show that VAR is the appropriate measure upon which to build optimal decision rules. VAR does not measure "event" (for example, market crash) risk. That is why portfolio stress tests are recommended to supplement VAR. VAR does not readily capture liquidity differences among instruments. That is why limits on both tenors and option greeks are still useful. VAR doesn't readily capture model risks, which is why model reserves are also necessary.

Because VAR does not capture all relevant information about market risk, its best use is as a tool in the hands of a good risk manager. Nevertheless, VAR is a very promising tool—one that will continue to evolve rapidly because of the intense interest in it by practitioners, regulators, and academics.

Barry Schachter is Vice President, Market Risk Portfolio Manager at Chase Manhattan Bank. He maintains a web page of VAR resources at http://pw2.netcom.com/~bschacht/varbiblio.html.

Seven Quantitative Insights into Active Management—Part 5

Data Mining Is Easy

by Ronald N. Kahn

Editor's Note: *The following article originally appeared in the Winter 1998 issue of the Horizon, a publication of BARRA, Inc., and is reprinted with permission.*

Why is it that so many strategies look great in backtests and disappoint upon implementation? Backtesters always have 95% confidence in their results, so why are investors disappointed far more than 5% of the time? It turns out to be surprisingly easy to search through historical data and find patterns that don't really exist.

To understand why data mining is easy, we must first understand the statistics of coincidence. Let's begin with some noninvestment examples. Then we will move on to investment research.

The Statistics of Coincidence

Several years ago Evelyn Adams won the New Jersey state lottery twice in four months. Newspapers put the odds of that happening at 17 trillion to 1, an incredibly improbable event. A few months later, two Harvard statisticians, Percy Diaconis and Frederick Mosteller, showed that a double win in the lottery is not a particularly improbable event. They estimated the odds at 30 to 1. What explains the enormous discrepancy in these two probabilities?

It turns out that the odds of Evelyn Adams winning the lottery twice are in fact 17 trillion to 1. But that result is presumably of interest only to her immediate family. The odds of someone, somewhere, winning two lotteries— given the millions of people entering lotteries every day—are only 30 to 1. If it wasn't Evelyn Adams, it could have been someone else.

Coincidences appear improbable only when viewed from a narrow perspective. When viewed from the correct (broad) perspective, coincidences are no longer so improbable. Let's consider another noninvestment example: Norman Bloom, arguably the world's greatest data miner.

Norman died a few years ago in the midst of his quest to prove the existence of God through baseball statistics and the Dow Jones average. He argued that "BOTH INSTRUMENTS are in effect GREAT LABORATORY EXPERIMENTS wherein GREAT AMOUNTS OF RECORDED DATA ARE COLLECTED AND PUBLISHED" (capitalization Bloom's). As but one example of thousands of his analyses of baseball, he argues that the fact that George Brett, the Kansas City third baseman, hit his third home run in the third game of the playoffs, to tie the game at 3–3, could not be a coincidence—it must prove the existence of God. In the investment arena, he argued that the Dow's 13 crossings of the 1,000 line in 1976 mirrored the 13 colonies which united in 1776—which also could not be a coincidence. (He pointed out, too, that the 12th crossing occurred on his birthday, deftly combining message and messenger.) He never took into account the enormous volume of data—in fact, an entire New York Public Library's worth—he searched through to find these coincidences. His focus was narrow, not broad.

With Norman's passing, the title of world's greatest living data miner has been left open. Recently, however, Michael Drosnin, author of *The Bible Code*, seems to have filled it.

The importance of perspective to understanding the statistics of coincidence was perhaps best summarized by, of all people, Marcel Proust—who often showed keen mathematical intuition:

> The number of pawns on the human chessboard being less

continued on page 19, column 1

Seven Quantitative Insights
continued from page 18

than the number of combinations that they are capable of forming, in a theater from which all the people we know and might have expected to find are absent, there turns up one whom we never imagined that we should see again and who appears so opportunely that the coincidence seems to us providential, although, no doubt, some other coincidence would have occurred in its stead had we not been in that place but in some other, where other desires would have been born and another old acquaintance forthcoming to help us satisfy them. (*The Guermantes Way, Cities of the Plain*, Volume 2 of translation of Marcel Proust's *Remembrance of Things Past* [New York: Vintage Books, 1982], p. 178.)

Investment Research

Investment research involves exactly the same statistics and the same issues of perspective. The typical investment data mining example involves *t*-statistics gathered from backtesting strategies. The narrow perspective says, "After 19 false starts, this 20th investment strategy finally works. It has a *t*-statistic of 2."

But the broad perspective on this situation is quite different. In fact, given 20 information-less strategies, the probability of finding at least one with a *t*-statistic of 2 is 64%. The narrow perspective substantially inflates our confidence in the results. When viewed from the proper perspective, confidence in the results lowers accordingly.

Four Guidelines for Backtesting Integrity

Given that data mining is easy, how can we safeguard against it? Here are four guidelines for data mining integrity:

- Intuition
- Restraint
- Sensibility
- Out-of-sample testing.

The *intuition* guideline demands that researchers investigate only those strategies with some *ex ante* expectation of success. Investment research should never involve free-ranging searches for patterns without regard for intuition.

The *restraint* guideline attempts to minimize the number of strategies investigated—that is, to keep the broad and narrow focus similar. In the best case, researchers decide *ex ante* exactly which strategies and variants they will investigate, run their tests, and look at the answers. They do not go back and continually refine their investigations.

The *sensibility* guideline deletes results that seem improbably successful. Observed, *t*-statistics that are too large may signal database errors or an improper methodology rather than a new strategy.

The fourth guideline, *out-of-sample testing*, is the statistician's answer to the curse of data mining. Coincidences observed over one dataset are quite unlikely to reoccur in another independent dataset.

Conclusions

Many backtesting results are not foolproof demonstrations of strategy value but merely coincidence. Four backtesting guidelines can help avoid data mining.

Ronald N. Kahn is Vice President and Director of Research at BARRA in Berkeley, California.

Integrated Approaches to Risk Management in the Financial Services Industry—A Seminar

December 8–9, 1997
Atlanta, Georgia

Anna Rappaport

This program, held at Georgia State University, was very significant in helping us to advance the work of risk measurement. All present gained by having dialogue with a diverse group of professionals, both in and outside the actuarial profession. We reviewed value at risk as well as other approaches for risk measurement.

The program planning arose out of our discussion of "gap analysis." In the 1996–97 Strategic Planning Committee, we focused on what needed to be done to make the mission and vision of the actuarial profession become a reality. The discussion of that issue, applied to the area of finance and investment, led to a discussion of value at risk. Planning Committee members Cindy Forbes and Irwin Vanderhoof then determined how to work on closing the gap, leading to a call for papers and then the seminar.

This program was truly a team effort sponsored by the Finance Practice Area and the Investment Section with leadership from Cindy Forbes, who heads our Finance and Investment Practice Area, and Irwin Vanderhoof. Other members of the Project Oversight Group were John Aquino, Harry Panjer, Bill Panning, and Jim Tolliver. Sheri Abel, Jackie Bitowt and Zain Mohey-Deen provided staff support.

As the financial services industry changes, the work of actuaries changes. Key changes include new approaches to the handling and management of risk on both the asset and liability sides of the balance sheet. The traditional disciplines of actuaries and other financial

continued on page 20, column 1