VAR and the unreal world

Richard Hoppe shows how the assumptions behind the statistical methods used to calculate VAR do not hold up in the real world of risk management

Value-at-risk is rapidly becoming the preferred means of measuring risk. But blindly accepting the assumptions that underpin its statistical methods can have adverse consequences for real risk managers operating in real markets. The purpose of this paper is to demonstrate that variance-based statistical methods are variably unreliable and that this unreliability is related to sample size in a counter-intuitive manner, to holding period and, possibly, to asset class. However, this is not a statistical article, it is an article about statistics.[1] Statisticians will not find elaborate derivations of equations or mathematical proofs here.

The reliability of risk estimates has its origins in psychometrics, the discipline associated with psychology that is concerned with the properties of tests and measurements. There are two principal properties in a psychological test: validity and reliability. Validity is whether a test actually measures what it purports to measure. Reliability is how consistently a test measures whatever it measures.

In modern portfolio theory, as well as in VAR applications, risk is defined as the volatility of returns. In turn, the volatility of returns is usually measured by the standard deviation of returns. Following the psychometric model, one can ask about the validity of the standard deviation as a measure of risk and about its reliability. How consistently do standard deviations and estimates based on them measure whatever is measured? Under what conditions can they be trusted?

There are well-developed technologies for assessing the reliability of psychological tests, but they are much more elaborate than is necessary for this paper. More important, they depend on the very assumptions at issue here. The issue of reliability in risk estimation does not require much statistical power to be explored.

Statistical assumptions

The use of measures such as standard deviation depends upon assumptions about the nature of the data being measured. If the assumptions are met, use of the measures may be unproblematic. If they are not met, there may be problems interpreting and using the numbers. It is therefore useful to review both the assumptions and the use of measures such as standard deviation in risk estimation.

Several assumptions underpin the use of linear, variance-based statistics to describe the dispersion (volatility) of distributions of market returns and the use of product-moment correlations to describe the relationship between a pair of time series of market returns. The principal assumptions are that:

- market returns are normally and independently distributed (NID); and
- the distribution of returns is stationary – as one moves through time, the mean and variance of the distribution are constant.

Product-moment correlations make the further assumption that only linear relationships between markets are of interest.

Research has shown that the assumption that distributions of raw market returns are NID is false. Though there is variation from market to market, distributions of daily returns of financial markets are generally both sharp-peaked and fat-tailed. In addition, some return distributions are skewed and, in the short term at least, there is evidence of serial dependence in some markets. Nevertheless, with some judicious massaging of the data – eg, detrending and using log returns rather than raw returns – and with a good deal of confidence in the robustness of linear, variance-based statistics in the face of violations at the extremes, standard deviations and product-moment correlations of historical returns are used for VAR estimation.

The attraction of standard deviation is that the properties of the normal distribution are very well understood. The reason a time series of market returns is forced into a NID distribution is to justify bringing linear variance-based statistical methods and probabilities to bear on risk estimation. The *RiskMetrics Technical Manual* says: "An important advantage of assuming that changes in asset prices are distributed normally [in spite of knowing that they are not] is that we can make predictions about what we expect to occur." The distribution is not defined by the data; it is chosen for no better reason than that we have some statistical tools available.

If the mean and standard deviation of a normal distribution are known, very precise probability statements can be made about the location of values in that distribution. For example, one can confidently assert that the probability that a randomly selected value will be more than +/–1.645 standard deviations away from the mean is 0.10, with half of the probability (0.05) in each tail. The ability to make such probability statements with high confidence is the property of normal distributions that VAR estimates depend on. That property depends on the robustness of the probability statements in the face of violations of the assumptions. VAR estimates are especially dependent on robustness with respect to violations of the assumptions in the tails of the distribution. Since VAR estimates are typically concerned with extreme probabilities, say, 0.05 or even 0.01, the question is not: how robust is the standard deviation with respect to violations of the assumptions in general? Rather, it is: how robust is it at the extremes in the face of violations of the assumptions?

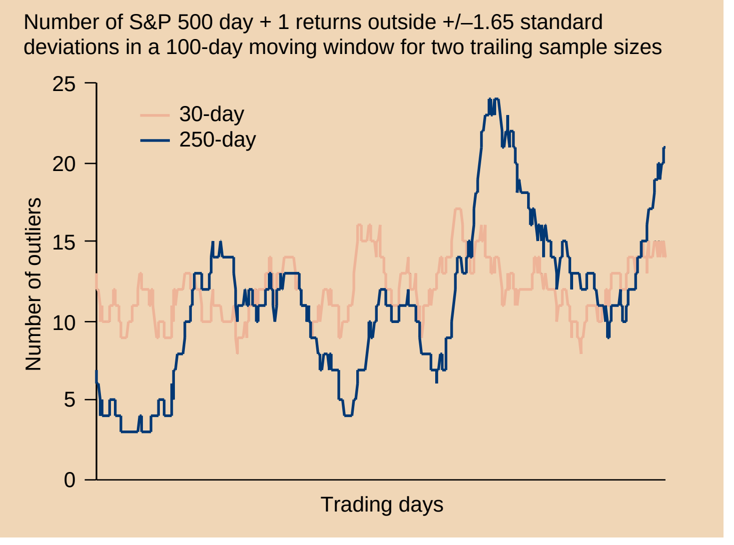
The short answer to the latter question is "not very". Here I do not intend to provide extended tests of the robustness of the standard deviation as a measure of the probability of extreme returns. Rather, my point can be made much more simply and directly by demonstrating a counter-intuitive property of the standard deviation as a measure of the probability of extreme market returns.

Reliability of real standard deviations

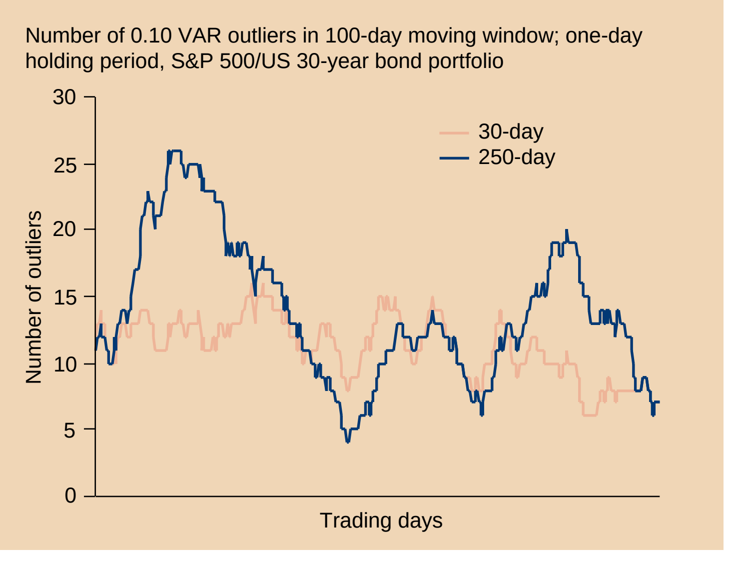
If I am a portfolio manager or bank officer concerned with risk control, I am not interested in the population parameters of a hypothetical distribution of an infinite population of returns. I am interested in the best possible answer to a simple question: "Under current conditions, what is the best possible estimate of my risk today?" The key phrase in that sentence

[1] *We will use the basic methodology described in JP Morgan's* RiskMetrics Technical Manual, *fourth edition, 1997*

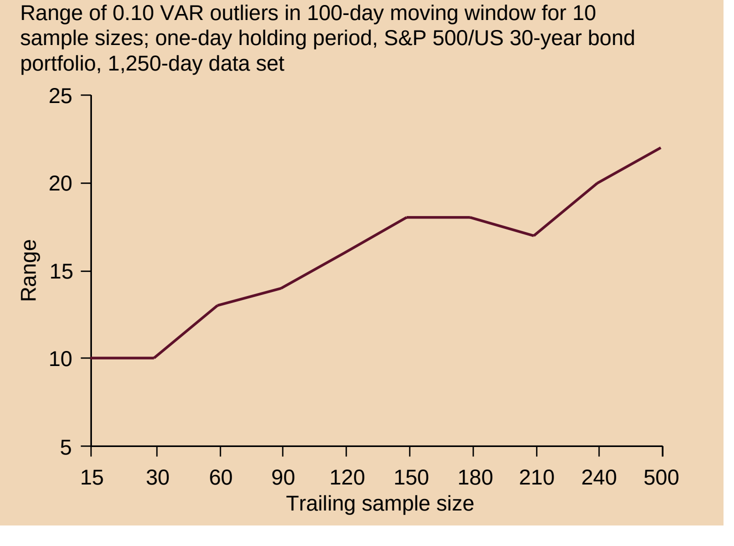
1. S&P 500 outliers for 30- and 250-day samples



2. VAR outliers for 30- and 250-day samples



3. Range of outliers v. sample size



is "best possible estimate". Take the assertion that the risk of losing a specified number of dollars or more in the next 24 hours is 0.05, a typical answer provided by variance-based estimation methods. The next question should be, "How good is that estimate? If I get 100 such estimates in the next 100 business days, how often and by how much will the actual number of losses of the specified size vary from the 0.05 asserted?"

The "how good is that estimate?" question is typically answered by appealing to somewhat circular theoretical arguments ("since we've assumed a normal distribution because the distribution of market returns passed our tests for normality, it's our best estimate because we know the properties of normal distributions") and/or by citing data showing that the proportion of outliers in the sample as a whole does not differ markedly from that expected given the normality assumption. For example, the *RiskMetrics Technical Manual* reports the proportion of actual outliers against the expected proportion in large samples for a number of the markets it analyses (the smallest sample reported is two years of daily data, roughly 500 trading days). But it does not report on the proportions of observations exceeding the limits tested in the sense of "reliability" as defined above, ie, the consistency with which a test measures whatever it measures through time. By summarising over entire data sets, the reported data throw away time, and time is important to real risk managers.

Consider the following exercise. Given a five-year (1,250 trading days) time series of daily returns from some market, say the S&P 500, suppose that one calculates a parallel series of moving standard deviations from a trailing sample of returns. Each day, one calculates the standard deviation for the preceding N days, and then looks at the next day (day 1) to see if the return at the close of trading on day 1 is outside +/–1.65 standard deviations. Then one creates a data series of 900 values consisting of the numbers of such outliers in a moving window of 100 days depth.

Now, one has a choice of a 30- or 250-day trailing sample from which to calculate the standard deviation. Which will provide the most reliable results, in the sense that the proportion of day 1 outliers in the 100-day moving window is least variable over the most recent 900 days?

Figure 1 contains the answer to the question for the S&P 500. The more reliable estimate of risk is provided by the 30-day trailing standard deviation, where "more reliable" means that the actual number of outliers in the 100-day moving window is less variable through time for the 30-day sample than for the 250-day sample.

Interestingly, in informally trying this exercise with various people, those who are statistically sophisticated tend to get it wrong, while those who are statistically untutored tend to get it right. There is a good reason for this. In introductory statistics, people are taught the law of large numbers and the central limit theorem. Large samples are better than small samples. The larger the sample, the more nearly normal the sampling distribution and the smaller the standard error of estimate of a population parameter from a sample statistic. Therefore, the 250-day sample should give a better estimate of the population parameter than the 30-day sample and therefore (here comes the leap) the 250-day trailing sample should be more reliable.

In orthodox sampling theory, that is correct; large samples are better than small samples. (I am ignoring issues associated with the relations between statistical power, effect size and Type I error level.) However, in the methods used to estimate near-term risk in financial markets, the law of large numbers and the central limit theorem are dangerously deceptive. In estimating near-term risk, one is wholly uninterested in population parameters. One is interested only in the likely state of affairs tomorrow or next week. A hypothetical infinitely large and normally distributed population of market returns is at best irrelevant to that problem; at worst, it is actively misleading. The statistically untutored respondents explain their choice of the 30-day sample by saying things such as: "Well, what happened a year ago isn't relevant; what happened recently is what's important." This is the same reasoning that led JP Morgan to use 74-day exponential weighting as the basis for the RiskMetrics data set, and it is correct.

The reverse is true for random samples of returns, as orthodox sampling theory suggests. If samples of 30 and 250 are randomly selected without replacement from the set of 1,250 days, the larger sample provides more reliable "forecasts" for the 1,250 returns. However, a 250-day random sample is not as reliable as a 30-day trailing sample, even for the

A. Ranges and median numbers of outliers in several VAR analyses

Market	Sample size	FTSE 100				S&P 500				US bond				UK gilt				Bund				Sfr FX				DM FX			
S&P 500	30	14	22	11	18	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	250	13	26	9	8	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
US bond	30	12	24	10	14	10	37	11	16	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	250	19	34	8	9	22	48	13	19	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
UK gilt	30	14	25	10	11	13	32	11	16	15	30	9	14	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	250	26	29	5	4	21	48	10	13	27	52	9	12	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Bund	30	14	31	10	13	11	28	11	17	10	30	12	14	16	29	11	15	–	–	–	–	–	–	–	–	–	–	–	–
	250	20	22	5	2	16	31	9	10	22	29	8	12	29	48	6	7	–	–	–	–	–	–	–	–	–	–	–	–
Sfr/$ FX	30	9	28	10	12	10	30	11	18	12	26	11	16	11	22	12	15	11	30	11	16	–	–	–	–	–	–	–	–
	250	12	24	8	11	17	40	9	18	15	38	11	11	14	33	9	15	19	44	13	18	–	–	–	–	–	–	–	–
DM/$ FX	30	9	27	10	10	10	28	11	17	12	25	10	16	12	22	11	15	13	30	10	15	13	43	10	21	–	–	–	–
	250	13	22	7	8	17	42	9	18	21	39	9	14	19	30	10	13	15	44	11	18	19	44	6	12	–	–	–	–
¥/$ FX	30	11	39	11	10	9	41	10	17	9	40	10	18	10	33	10	13	11	38	10	15	11	30	9	16	13	24	10	15
	250	15	31	7	6	15	43	9	15	20	53	8	10	20	42	9	14	20	51	12	18	23	46	8	9	19	49	6	13

1,250-day population from which the 250 values were randomly selected. The 250-day random sample's range of outliers (maximum minus minimum) in a 100-day moving window is twice the range of the 30-day trailing sample. Note that for the more reliable 30-day trailing sample, all forecasts are for novel days, while the 250-day random sample is attempting to "forecast" the same population from which it was drawn, including the 250 days that were used to calculate the standard deviation. An implication of the reversal of the expected sample size effect for trailing samples of market returns is that either the distribution of returns is not stationary or returns are not serially independent, or both. As a consequence, standard sampling theory and the statistics that depend on it are inapplicable to risk estimation.

The reliability of VAR estimates

To address this issue in a slightly more complex situation, I tested VAR estimates for a two-component portfolio composed of equal dollar-long positions in the S&P 500 and the 30-year US Treasury bond, using the daily near-month futures price series from January 1, 1991 to October 11, 1996 as the basic data set. Two trailing sample sizes – 30 days and 250 days – were used initially to calculate the standard deviations of returns of the two markets and the correlation between the markets, the two inputs to VAR estimates. The one-day, 0.10 two-tailed VAR was calculated for each day, following the methodology described in the *RiskMetrics Technical Manual*, with two exceptions: I used unweighted values rather than exponential weightings, and I used log returns rather than raw percentage returns.

As in the standard deviation exercise above, I counted the number of outliers – occasions on which the actual day 1 return was larger or smaller than the VAR estimated at the 0.10 level – in a 100-day moving window for both series of estimates of daily VAR. Figure 2 shows the relevant results. As for the standard deviation, the shorter trailing sample produced more reliable VAR estimates than the longer one.

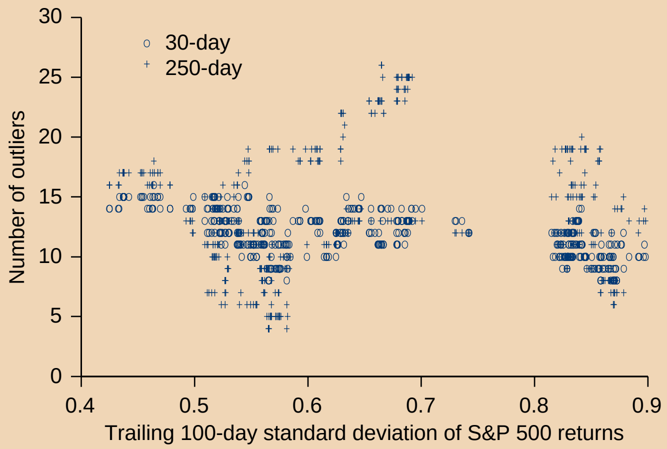
It transpires that the reliability of one-day VAR estimates is not linear or even monotonic in sample size. Repeating the VAR analysis described in the preceding paragraph, using sample sizes from 15 to 500 trailing days, one finds that reliability, measured by the range (maximum minus minimum) of the number of outliers in the 100-day moving window, displays an inflected curve across sample sizes, with the inflection occurring between samples of 150 and 240 days. Figure 3 shows the relevant results. Note that there is a floor effect: the minimum number of outliers is zero, compressing the range on the low end, so interpreting the curve shown on figure 3 is not straightforward. The non-monotonicity might also be an artefact of the particular period used for the test. What is clear, however, is that using a year of trailing prices to estimate one-day VAR produces a range of actual outliers that is much wider than the range for a 30-day sample.

I repeated the two-component portfolio VAR analysis for all 28 of the pairs of eight financial markets (two equities, three interest rates, three currencies) over a one-day horizon. The blue figures in table A show the range of one-day outliers (maximum minus minimum) for each 100 days for 30- and 250-day trailing samples for the 28 pairs of markets. As is obvious, the 30-day sample produced a narrower range with just one exception, the FTSE 100/S&P 500 pair.

The main difference in the larger sample size is an increase in the maximum number of outliers per 100 days, which is important to risk managers. The larger sample size shows substantially more outliers during some peri-

4. VAR reliability v. standard deviations of returns

Number of outliers per 100 days for 0.10 VAR of S&P 500/US 30-year bond portfolio as a function of trailing 100-day S&P 500 standard deviation of returns



ods than the smaller sample. In other words, using longer trailing samples for VAR estimates produces periods in which there are more surprises. Lest one believe that most of the surprises are pleasant ones, the 250-day sample for the S&P 500/US bond pair in a 100-day period produces a maximum of 17 one-day losses that are greater than that forecast by the 0.05 negative VAR, while the 30-day sample produces a maximum of just nine such losses in a 100-day period. (The "normal" expectation is five.) The same relationship with sample size that characterises all outliers shown in figure 2 holds for negative outliers: the longer the trailing sample, the more unpleasant surprises one gets during some periods, up to more than three times the frequency expected on the NID assumption. A promise of probabilistic safety in the long run is worthless if one goes broke in the short run.

While I am wary of generalising too far based on the relatively limited data sets I have tested, there is a suggestion in the S&P 500/US bond VAR data that is consistent with the statistically untutored reason for choosing a shorter trailing sample. Figure 4 shows an X-Y plot of VAR reliability against the trailing standard deviation of returns of the S&P 500 for the 30-day and 250-day sample sizes. It suggests that the superior reliability of the 30-day sample for the S&P 500/US 30-year bond pair is most apparent during periods of high stock market volatility, while the reverse appears to be the case during periods of low volatility. The former backs JP Morgan's reasons for selecting a relatively short exponentially weighted sample, rather than the 250-day unweighted sample required by the Basle Committee on Banking Supervision. It may be that it is during market periods when one most needs reliable risk estimates that long trailing samples provide the least reliability. I emphasise that this is tentative; it is based on a limited data set and more research is needed.

Basle on VAR estimation

The Basle Committee has issued guidelines for the use of VAR estimates by regulated banks. For the purposes of this discussion, the three relevant guidelines are that VAR estimates must:

- be based on a sample size of at least one year of data (roughly 250 trading days) or a weighted sample with an average lag of no less than six months;
- use a 0.01 probability value (ninety-ninth percentile, one-tailed confidence interval); and
- estimate the risk for a 10-day holding period.

To evaluate these guidelines in the light of the findings reported above, I repeated the two-sample VAR tests using a 10-day horizon rather than a one-day horizon. The Basle Committee allows VAR estimates calculated for a shorter time interval to be scaled up by a factor equivalent to the square root of time, so I used the one-day VAR estimates calculated above, multiplied by the square root of 10. The Basle Committee's decision to allow VAR estimates to be scaled as the square root of time follows from the assumption that returns are NID and serially independent.

As above, I counted the number of outliers in a 100-day moving window. For this exercise, however, "outlier" has a slightly different meaning. Instead of a one-day, close-to-close excursion greater than the VAR estimate, an outlier is now defined as the largest day-0-close-to-day-N-close excursion within the 10-day period, where N varies from one to 10. For example, take a two-asset portfolio long both assets, where both asset prices move down sharply during the first five days of the 10-day holding period and then move back up so that there is little or no change from day 0 close to day 10 close. The sharp day 0 to day 5 excursion may have gone outside the confidence limit defined by the VAR estimate and thus will be counted as an outlier even though the day-0-close-to-day-10-close return is inside the VAR limit. The day-0-to-day-5-excursion is, after all, a loss greater than expected on a daily mark-to-market basis. I repeated this exercise for the 28 pairs of markets. The red figures in table A show the ranges of the number of "within 10-day outliers" for the portfolios and sample sizes.

The pattern of results of this exercise essentially mirror those of the one-day holding period. As the table shows, for 23 of the 28 pairs, the 30-day trailing sample is more reliable, while for five pairs the 250-day sample is more reliable. Four of the five pairs in which the 250-day sample is more reliable have the FTSE 100 as a component, as did the sole exception in the one-day data. What seems clear is that the reliability of VAR estimates depends on interactions among holding period, sample size and (possibly) unidentified characteristics of the particular assets or asset classes contained in a portfolio. At the least, this raises questions about a "one size fits all" approach to VAR estimation.

Finally, the relationship between sample size and reliability of VAR estimates holds for longer time series. I tested VAR estimates for all 15 pairs of an equity market, interest rate market and four foreign exchange rates in a 3,800-day data set for a one-day holding period. The differences reported above between the 30-day and 250-day samples in estimating one-day VAR are slightly attenuated but still very clearly present in all 15 pairs. The relationship between trailing sample size and reliability is not an artefact of the period chosen for test. Table B shows the ranges of 100-day VAR outliers for the 15 pairs in the 3,800-day data set.

Summary of major findings

There appears to be an interaction among trailing sample length and holding period along with a possible third variable, portfolio composition. Four main results are evident:

- Down to some lower limit, shorter trailing samples usually produce more reliable VAR estimates than longer ones. This is true for one-day and 10-day holding periods.
- Across sample length and asset class, VAR estimates for the one-day holding period are consistently and substantially more reliable than for the 10-day holding period.

B. Range of number of VAR outliers in 100-day moving window for 15 pairs of markets

Market	Sample size	S&P 500	US	Sfr/$	£/$	DM/$
US	30	16	–	–	–	–
	250	27	–	–	–	–
Sfr/$	30	15	13	–	–	–
	250	23	17	–	–	–
£/$	30	11	12	18	–	–
	250	23	24	25	–	–
DM/$	30	13	12	14	18	–
	250	21	20	25	26	–
¥/$	30	15	14	15	19	14
	250	25	24	25	22	25

Note: one-day holding period; 3,800-day data set

☐ A market or asset class effect may be associated with the FTSE 100, with pairs involving the FTSE producing five of the six reversals of the general sample size effect.

☐ The greater reliability of shorter trailing samples holds for long time series; it is not an artefact of a particular period or market regime.

Implications

What does one make of all this? First, these findings imply that asserting a VAR probability estimate with two-decimal-place precision at the 0.10, 0.05 or 0.01 level seriously misrepresents the precision possible regardless of sample size, holding period or asset class. The apparent exactness of the probability statement can mask more than an order of magnitude of variation in the actual probability of loss on a time scale appropriate to the practical situation of a risk manager – months and quarters rather than decades. For an S&P 500/US bond portfolio and a 250-day trailing sample, on any given day the probability (measured as the frequency of occurrence per 100 days) of incurring a one-day loss to a long position greater than that specified by the VAR estimate may be 0.05, 0.17 or 0.00. For a 30-day trailing sample, the range of variation is narrower, but it is not trivial. The strongest statement one can honestly make is that the probability of a loss of the specified magnitude at the calculated ninety-fifth percentile is in the fuzzy neighbourhood of 0.05-ish. That is no doubt unsatisfying to a risk manager or regulator, but to pretend otherwise is to mislead oneself and one's clients. The putative precision of a VAR probability estimate with two significant digits to the right of the decimal point is deceptive.

Second, within broad limits, for risk estimation, shorter samples can be substantially more reliable than long samples. In this vein, recall that the Basle Committee has opted for a one-year unweighted trailing sample and a 0.01 one-tailed probability as the standard for VAR estimates. As figure 3 and table A show, the committee is erring far out on the large sample side, thereby guaranteeing substantially less reliable estimates than is possible. The RiskMetrics data set uses 74-day exponential weighting to estimate VAR. I have not tested 74-day exponentially weighted data, but my bet is that they are similar to the 30-day unweighted data. Given the FTSE results above, both the Basle requirements and the RiskMetrics data set are subject to the "one size fits all" question.

Third, use of the broad array of modern statistical methods without a clear understanding of the implications of their assumptions for the actual real-world application to be modelled always needs examining. The mathematical sophistication and complexity of the techniques can mask a deep misconception of the applied problem. In managing risk, one is interested in as dependable an answer as possible to the question "What is my risk?" Given a distribution of returns that is non-normal, especially at the extremes, and probably also non-stationary and/or serially dependent, the seeming exactness and "scientific" appearance of variance-based estimates of risk misrepresent the real situation. The alleged precision is far beyond what is possible. Paradoxically, risk managers might often be better off depending on weaker, small-sample, non-parametric estimation methods.

The perceptive reader will have noticed that I have not mentioned the actual proportions of outliers in any of the results reported above. That is, I have not reported the performance of the various conditions with respect to the number of outliers over the whole time series. The reason is simple: in designing a risk estimation system, one's first interest should be reliability. Once one has devised a reliable means of estimating risk, one can proceed to tune the system to achieve the confidence limits desired if they are possible. It is best to take as much as the data are capable of producing, but not to torture the data beyond what they can tolerate. The green (one-day holding period) and black (10-day holding period) figures in table A show the median number of outliers for the 28 pairs for the two holding periods for the shorter data set. As may be seen, for the one-day holding period the median number of outliers hovers close to 10, the expected value. For the 10-day holding period, the medians are more variable and tend to be higher than the expected value, with the smaller sample size showing greater departures from the expected value of 10. I report medians rather than means because the distributions of outliers are severely compressed on one boundary – they cannot go below zero – and so means are misleading representations of the central tendency of the distribution. I report no significant digits to the right of the decimal point because, as I have argued above, that level of precision is at best misleading.

Alternatives

Given that variance-based risk estimates in general and VAR estimates in particular are variably unreliable, what does one do? First, one must recognise the fact of unreliability. The array of powerful statistical techniques available to the risk manager – to the extent that they depend on the assumptions of normality at the extremes, serial independence and stationarity – are founded on quicksand. Explicit recognition of unreliability moves the focus from massaging numbers in ever more complex ways to devising defences against risk in the face of uncertainty in the Keynesian sense of the word. Depending on unreliable estimates can be costly. To design appropriate defences, risk managers need to know just how unreliable their estimates are.

An alternative is to try to devise risk estimation techniques that avoid or mitigate the problems inherent in linear variance-based statistics. My company has been exploring alternative risk estimation techniques. We use proprietary time series representation and pattern recognition algorithms that produce descriptions of the patterns of linear and non-linear relationships within a cluster of related markets through time. Given a historical database of such descriptions and given the description of the pattern for today, our programs select instances from the past that display patterns most similar to today's pattern. We use non-parametric measures of the dispersion – percentiles – of that selected sample to estimate today's risk. For comparable sample sizes, this approach often (though not always) produces more reliable (in the sense defined earlier) risk estimates.

> The array of powerful statistical techniques available to the risk manager... are founded on quicksand

The most reliable risk estimates we have so far produced are created by a hybrid of the two approaches. Using both the VAR calculated from an unweighted trailing sample and the non-parametric risk estimates from samples selected by our technology, and simply taking the larger of the two on each day produces improvement in the empirical reliability of risk estimates. For example, in the 250-day sample case, for the S&P 500/US bond pair the range of variation in the number of negative outliers for each 100 days is reduced from 17 to 10, the reduction being accompanied by an increase in accuracy as measured by the difference between the total number of outliers and the expected value. The reduction in the range of variation is wholly accounted for by a decrease in the maximum number of outliers per 100 days.

We have not yet completed all the research necessary to evaluate the hybrid methodology further. A central part of the problem, of course, lies in the way one represents the patterns of interactions among markets and how one defines "similarity".

This is not meant to claim that our approach is the best possible way to estimate risk, though naturally we are attracted to it. However, it is meant to demonstrate that there are alternative ways to approach risk estimation aside from (or in addition to) linear variance-based statistics. Market returns are neither NID nor stationary, and regardless of how one massages the data, violations of those assumptions can lead to serious practical consequences. Nevertheless, the past can tell us about the future provided that we know which properties of the past we should pay attention to and how we should interrogate the past about those properties. The fundamental point is this: believing a spuriously precise estimate of risk is worse than admitting the irreducible unreliability of one's estimate. False certainty is more dangerous than acknowledged ignorance. ■

Richard Hoppe is chief executive officer of IntelliTrade, a risk management decision support firm
e-mail: itrac@itrac.com