How historical simulation made me lazy

Christopher C. Finger
chris.finger@riskmetrics.com

April 2006

When I started working in risk management, about ten years ago, we argued a lot, as researchers anywhere do. Noticeably, what we argued about was forecasting, particularly about which volatility estimates worked better, and which distributions captured returns well. At the same time, we operated under the restrictions of transparency and model parsimony: since we were trying to build risk models for extremely diverse portfolios, it would not do to have a different parameterization for every security, or even asset class.

Eventually, we became comfortable with our current volatility scheme and with the assumption that given the volatility, returns are normally distributed. We have not changed this much since, for both good and bad reasons. A good reason is that this scheme holds up well, certainly compared to alternatives that are just as straightforward. A bad reason is that the market—and consequently, we—have stopped having so many arguments about forecasting.

A disturbing trend, which may be either a cause or an effect of our lack of arguments, is that many institutions have come to rely on historical simulation as their sole statistical measure of risk. Again, there are good and bad reasons for this, but what is most worrisome is that the reliance on this methodology represents just giving up, and failing to acknowledge that statistical forecasting, with all of its imperfec-

tions, is still informative.

In this piece, we examine the reasons most often cited for using historical simulation. Along the way, we will issue a few reminders of things that we know and should not throw away. We finish with a plea that the risk community not give up, but rather go back to arguing. It has been a bit too quiet on this front for too long.

The reasons

When risk managers are asked why they opt for historical simulation, they usually respond with one or more of the following:

1. It is easy to explain.

2. It is conservative.

3. It is "assumption-free".

4. It captures fat tails.

5. It gives me insight into what could go wrong.

As a result of the first two of these, and perhaps the third as well, there is another reason: "my regulator likes it."

Noticeably absent from the list of reasons is the statement

1. It produces good risk forecasts.

If we believe all of these statements, then the choice of historical simulation can be seen as the ultimate sacrifice of performance for transparency and simplicity. While we will argue that these statements are not necessarily true, it is also worth questioning whether we should accept a total victory by the transparency camp, rather than the tension that existed when we first started discussing the models.

Easy to explain?

A concise definition of historical simulation is as follows:

> *Apply the last one year's worth of daily price changes to my existing portfolio. Pick the 1% quantile of the portfolio profit-and-loss results. This is the VaR forecast.*

Simple enough, but there are still questions that remain, most of which we must ask for any statistical model:

What do we mean by price changes? Should we apply absolute differences or returns?

More crucially, what do we mean by prices? Do we examine prices of the precise securities that we hold, or some smaller, and possibly more meaningful, set of risk factors?

Finally, how do we obtain risk estimates for longer horizons? We have not made any assumptions about the evolution of risk factors. This leaves us with the choice of using (abusively) the "square root of time rule", or sampling historical returns over periods commensurate with the horizon, which quickly reduces the amount of data from which we can make statistical estimates.

So the methodology is in fact simple, but as with any forecasting model, there are details to fill in. We will return to the second of these questions in particular later.

Conservative?

That historical simulation is conservative is not entirely clear, since we effectively assume that the worst return of the last year is the worst thing that could happen.

Still, even if we grant that the model is conservative, this is hardly a criterion to select a model. For an extreme example, consider some of the backtesting results we presented in a note last year. Two of the banks we considered, JP Morgan Chase and Société Générale, disclosed in their annual reports that they used historical simulation to produce risk estimates. JP Morgan Chase disclosed just three days over the prior four years on which their realized loss exceeded their 99% VaR estimate; Société Générale disclosed that on no day in the prior three years had they experienced such a loss. The likelihood of either of these events, assuming that the VaR models do produce good forecasts, is well under 1%. Clearly, the disclosures are meant to demonstrate to shareholders that the banks risk numbers are conservative.

A savvy shareholder, however, should reason that if the banks are truly managing their capital based on these risk estimates, and they are never experiencing losses of the level that they should, then the banks are either overcapitalized or not taking enough risk;

both of these should worry the shareholder. The answer, we hope, is that the banks are not really managing themselves against the numbers in the annual report, and that they are at most guilty of useless disclosure, but not of inadequate risk appetite.

So there is conservative, and there is useless. In this case, unfortunately, the two go hand in hand.

Assumption-free?

This is the most frustrating description to hear. The fact that a model is so easy to state does not make it free of assumptions, but rather makes the assumptions less explicit. The danger here is that we get lulled into a state where we forget the assumptions we are making, and stop questioning them.

So to question assumptions, we should first state some. When we build a forecast based on historical simulation, we assume, among other things, that

1. The historical days on which we sample adequately represent the distribution of future returns on our portfolio,

2. We sample enough points to have a statistically significant estimate of the desired quantile of the distribution, and

3. The price data we choose to sample from is representative of our holdings.

These are really no different from the assumptions for any statistical risk model: we choose appropriate time series (3), we forecast those series well (1), and we make precise conclusions about portfolio risk (2).

For the first assumption to hold, the return distribution should be constant (or close enough) over the one-year historical period. There is much prior research to refute this notion; a veritable industry of volatility forecasting exists, and one of the pioneers in volatility forecasting, Robert Engle, was awarded the Nobel prize in 2003. Certainly, volatility forecasting would not have gained the attention and acceptance it has were return distributions constant over long periods.

To illustrate the implications of ignoring volatility fluctuations, we examine the realized volatility for four financial time series: the S&P 500 equity index, the USD/JPY exchange rate, the January 2006 crude oil futures contract, and a return series based on systematically selling one-month at-the-money puts on the Euro. The last series is intended to mimic a dynamic strategy, such as we might be exposed to as a hedge fund investor. We present the results in Figure 1. The realized volatility plotted is simply the standard deviation of daily returns over the previous month. We observe that volatility changes, moving by as much as fifty percent within a given year. And since volatility changes, there is a limit to how much historical data can be relevant for forecasting from today; we cannot simply improve our forecasts by including more history.

Observing that volatility fluctuates is only the beginning of our forecasting challenge, though. The more crucial question is whether there is anything we can do about it: is there information in the past that can help us forecast the future level of volatility? If not, then fluctuation in volatility is an interesting academic observation, but not an effect we can exploit in risk management. In Figure 2, we present the lagged correlation in return magnitude for the same

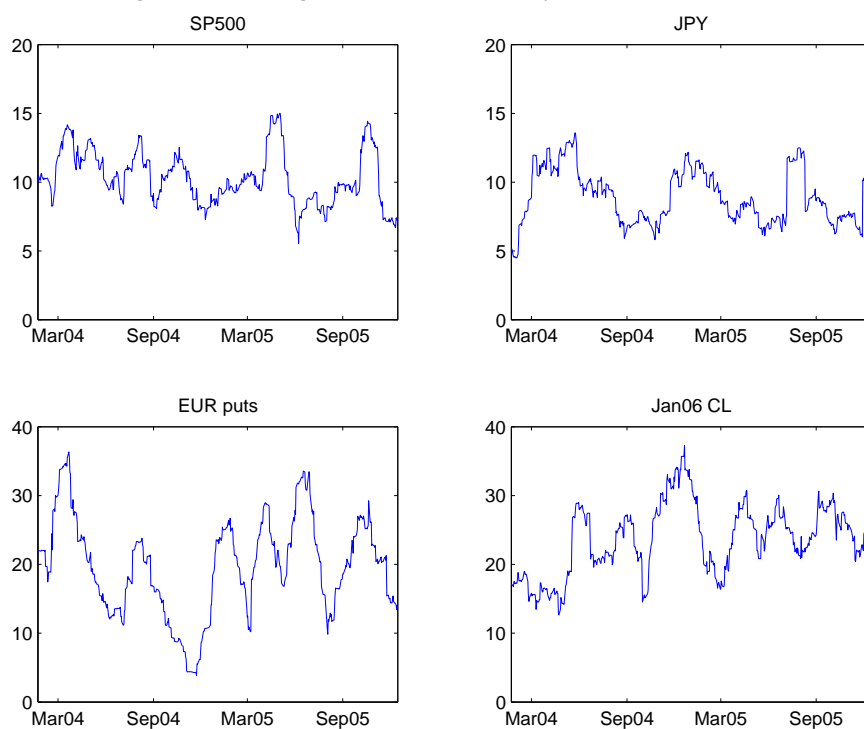Figure 1: Rolling one-month volatility (annualized, in %)

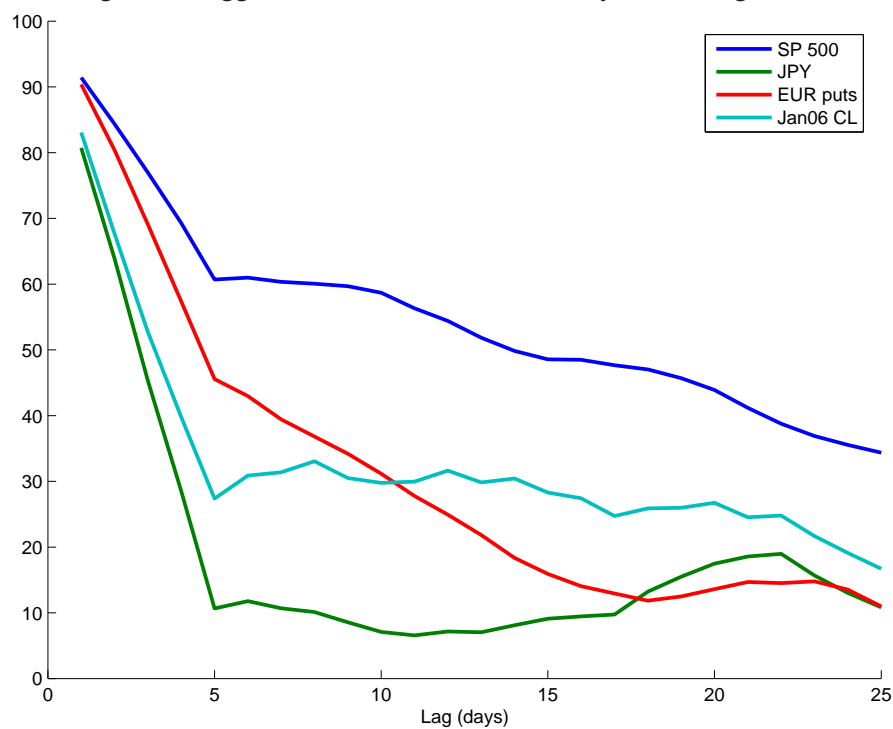Figure 2: Lagged correlation (in %) of weekly return magnitudes

Table 1: Historical simulation over 2005. Worst case losses (in %)

	VaR 99%		VaR 95%	
	Lower (5)	Upper (1)	Lower (17)	Upper (8)
S&P 500	1.41	1.68	1.06	1.29
JPY	1.11	1.43	0.84	0.99
EUR puts	3.94	4.65	2.44	3.23
Jan 06 Crude Oil	2.86	3.77	2.22	2.69
Near-month Crude Oil	3.69	4.90	2.78	3.55

four time series. For example, we calculate the correlation between the size of a return at one point in time with the size of the return on the same time series ten days into the future. From the figure, we see that significant correlations exist, meaning that information about the sizes of past returns can help us forecast the magnitudes of returns in the future. *How* we do this is something we should argue about, but that forecasting is possible is a fact we should not dismiss.

The second assumption is that the VaR produced by historical simulation has statistical significance. To assess this assumption, let us, despite evidence to the contrary, accept the first assumption: that the past 250 daily returns are in fact representative of the one true distribution. The question then is whether these points are sufficient to reliably estimate the tail of the distribution. In this case, the expected number of portfolio loss scenarios that exceed the true 99% VaR level is 2.5; our best estimate of VaR is thus somewhere between the second and third largest losses. However, there is roughly an 80% probability that between one and four scenarios fall below the true VaR level[1]; this implies that if we

want 80% certainty in our VaR estimate, we can only conclude that the true VaR level is between the first and fifth largest losses. Similarly, there is roughly an 80% probability that the 95% VaR level falls between the eighth and seventeenth largest losses.

We present the confidence intervals for a selection of time series in Table 1. The uncertainty in the VaR estimates, measured by the width of the band relative to the middle of the band, is significant: as high as 35%, and 25% on average. So if we interpret historical simulation as a statistical technique, there is still a significant amount of estimation error in the VaR forecasts that result. And if we ask more of the historical simulation process—for instance, higher confidence levels or the attribution of portfolio VaR to individual positions—the estimation errors are even greater. Finally, since our scenarios are not in fact independent, as we observed in Figure 2, the effective size of our sample is even fewer than the 250 points, meaning our estimation error in truth is worse than what we present here.

The third assumption relates to a theme we have treated many times: how to choose good risk factors. In our February note, we addressed futures

---

[1]This follows from the observation that if the scenarios are independent, then the number of scenarios that fall below the true VaR level follows the binomial distribution.

contracts. We made the point that for many commodities, there is a tendency for a futures contract to become more volatile as its delivery date becomes closer. This effect is strong enough that we must account for it when we forecast risk; either we use data on specific contracts, but incorporate the effect in our volatility model, or we use a simple model but choose a time series that is more representative of what might happen to our position from today. Historical simulation being at best a simple model, it is incumbent on us to choose our data wisely.

As an example, suppose we held the January 2006 crude oil contract at the end of 2005. To forecast the risk of this position with historical simulation, we have essentially two options: either we examine the historical returns on the actual January contract, or we examine returns on some form of constant maturity contract. For our example, we simply take the historical returns on whatever was the nearest contract at the time; the historical returns from April are those on the May contract, while those from December are on the January contract that we hold. We present the VaR estimates using the two approaches in the last two lines of Table 1. The differences are thirty to forty percent, and the statistical bands around the estimates only barely overlap.

Which of the two VaR forecasts is better? Our opinion is that the second is, but that is a topic for a different argument. For the matter at hand, the crucial point is that the two forecasts are different, and a risk manager must make an explicit choice between the two. The choice is equivalent to what we thought we were avoiding: an assumption.

After the bad and the ugly

Of the reasons to utilize historical simulation, we should now turn to the good.

The fourth reason we listed is that the methodology captures fat tails: that is, that very large returns occur more frequently than they do under the Gaussian distribution. While this is true, it is worth issuing one note of caution.

It is important to remember that fat tails can be simply volatility clusters in disguise; a time series with systematically varying volatility and Gaussian conditional returns can appear, if we do not recognize that the volatility is forecastable, to have fat tails. This is not to suggest that we abandon the use of fat-tailed distributions, but rather that we be sure to first forecast what we can, and restrict our application of alternative distributions to what is left over.

A reasonable application, then, is to first apply a good volatility forecasting method, then divide each observed return by the volatility that would have been forecast on that day, thereby creating a time series of *residuals*. If we have done well with our forecasting, then we should have extracted everything we can from the returns, leaving behind in the residuals only the true randomness, and nothing that we can predict. It is this distribution we should consider for fat tails. Sampling from the historical residuals is one method to capture these, though it still leaves us with the estimation error we discussed earlier. A more robust approach, then, may be to fit a standard distribution to residuals observed across many time series, from which we may draw arbitrarily many returns in a Monte Carlo process. This is another argument worth having.

While it is exciting to work with distributions other

than the Gaussian, we should look at the effect of this work in context. A realistic distribution for most financial time series is a t distribution with five degrees of freedom. At 99% confidence, the differences in VaR between the Gaussian distribution and this more realistic distribution[2] is 12.0%. Decreasing the degrees of freedom to three (producing even fatter tails) increases the difference to 12.7%. These are significant differences, but pale compared to the potential 50% and 35% discrepancies we risk by ignoring volatility fluctuations or estimation error.

The fifth reason is our favorite. Ultimately, the most appealing thing about historical simulation is that our scenarios have labels. It is one thing to say that a critical loss occurs in scenario number 347 of a Monte Carlo process; it is quite different to say that a critical loss occurs with a repeat of September 19. In the latter case, we gain more intuition about our risk than just knowing we are hurt by a rise in the five-year interest rate or JPY implied volatility. We can ask ourselves how concerned we are about a repeat of a specific day in history much more easily than we can assess the likelihood of a broad combination of moves across many markets.

This is not to say that we cannot gain insight from the scenarios in a Monte Carlo process. In fact, there has been discussion recently[3] encouraging risk managers to examine more closely those scenarios that produce the worst losses under Monte Carlo; and in fact, we discussed in a piece last year the benefits of the maximum loss technique in identifying bad portfolio scenarios. The examination of scenarios is useful, but the distillation of a scenario to a single piece of information—a date—is hard to top. Thus, beyond simply examining bad Monte Carlo scenarios,

we can attempt to label them. For a specific Monte Carlo scenario—which encompasses changes in all of our risk factors—we might ask which historical return, or which of a set of stress scenarios, is in a sense closest. Though we would not see a perfect match, we might still gain the intuition, for example, that we should worry more about a repeat of July 2 than of August 20 of last year, or more about events like the 1997 Asian crisis than the 2001 bear market. Here, in contrast to both historical simulation or stress testing, our worrisome scenarios would be a product of both how we are exposed and what scenarios are likely in the near future.

Final thoughts

Since there have been statistical risk models, there have been warnings to understand the model weaknesses. As we started to use the statistical models to forecast risk, the wise were there to urge us to "know what we don't know."

By abandoning proven statistical techniques in favor of the transparent, but flawed, forecasts of historical simulation, we seem to have answered that what we don't know is everything. Certainly, it is not prudent to rely too heavily on statistical forecasts, but it is also imprudent to not rely on them at all.

As we examined the reasons for using the historical simulation methodology, we found attractive features in the insight it can bring to bear on a portfolio. No one can argue that the last 250 days of historical returns are an interesting set of information. But the jump from information to forecast is a big one, and when historical simulation, or any

---

[2]Assuming the two distributions are scaled to have the same volatility
[3]See Rowe (2005).

statistical method, is used to forecast, it should be subjected to greater scrutiny: are the assumptions sound, and are the forecasts accurate? Under this scrutiny, the historical simulation method does not stand up.

One proposal, then, is that we stop thinking of historical simulation as an alternative to statistical forecasting techniques, and consider it instead as a complement to those techniques, in much the same way as we think about stress testing. Such a recategorization removes the burden of forecasting accuracy, and emphasizes the goal of providing insight. Judged under this different set of standards, historical simulation fares better than before.

A second proposal is that we take stock of the two most attractive features of historical simulation, and seek out ways to incorporate these into our statistical approaches. As discussed, we should revisit the historical return distribution, though only after we normalize those returns using an effective volatility forecast. And we should examine specific Monte Carlo return scenarios, and seek to align these with historical scenarios or even stress scenarios, in an effort to provide more intuition about what is likely to cause significant portfolio losses.

A last proposal is that we not be complacent. Market risk measurement is a mature field, but not a dead one. Statistical risk forecasts can be good, and can get better, but not if we remain satisfied with an inferior alternative.

Further reading

- Finger, C. (2005). What's the worst that could happen?, Research Monthly, March.

- Finger, C. (2005). Back to backtesting, Research Monthly, May.

- Finger, C. and O'Shea, L. (2006). Future risk, Research Monthly, February.

- Rowe, D. (2005). Whither stress testing? *Risk*, **18**(10): 65.

- Hull, J. and White, A. (1998). Incorporating volatility updating into Value-at-Risk calculations. *Journal of Derivatives*, **6**(1): 5–19.

- Pritsker, M. (2001). The hidden dangers of historical simulation. Finance and Economics Discussion Series, Federal Reserve Board.