What can you afford in a 1% world?

Louise Pryor asks whether we are testing our spreadsheets.

'Of course, in a 1% world we can't afford to test our spreadsheets properly', said the actuary I was chatting to. I was speechless. He was telling me a major UK life company could not afford to make sure its products were profitable, its business plans workable, and its financial reporting accurate. Surely I hadn't heard him correctly? But I had. His comment reflects a broad body of opinion in the actuarial world and elsewhere: that the short-term costs of testing spreadsheets outweigh the long-term benefits. I disagree, of course, and shall explain why.

Short-term costs

It's my impression that most people severely overestimate the short-term costs of testing. This is probably for a couple of reasons:

- they have little experience on which to base their estimates, because so little systematic testing is performed; and
- when they do try to do it properly they don't use the most efficient methods.

Without going into too much detail, there are many ways of going about testing, some of which are rather more effective than others. When times are difficult, for example when expense ratios are under pressure, it becomes much more important to use efficient methods in order to get the most bang for your buck.

Long-term benefits

The biggest misunderstanding about spreadsheet testing is in the area of long-term benefits. The principal benefit is that testing helps control the risk of your spreadsheets producing the wrong numbers. The principal misunderstanding concerns the magnitude of that risk. Think about it for a moment: what percentage of spreadsheets do you think contain errors? 10%, even 20% if you are a pessimist? Even those percentages are high enough to ring alarm bells, but all the evidence shows that the true proportions are many times higher.

Spreadsheet error statistics

A few years ago Professor Ray Panko, at the University of Hawaii, pulled together the available evidence from field audits of spreadsheets. His full report is available from his spreadsheet research page at *http://panko.cba.hawaii.edu/ssr/*; table 1 to the right summarises the results.

In total, 91% of the spreadsheets contained errors. It's difficult to be sure how representative these samples were, as it is quite possible the percentages are overestimated because only those spreadsheets thought likely to be problematic were audited. On the other hand, it is also possible that only those thought to be especially critical, and over which especial care had been taken, were included, which would lead to underestimation of the overall error rate. Another factor leading to underestimation is that these results are from reviews, not from testing. In general, the longer a review takes, the more errors are found; and we don't know how thorough these reviews were. Experiments indicate that, typically, about 50% to 80% of errors are found by reviews.

Looking at the percentage of individual cells with errors leads to roughly similar conclusions. Field audits show cell error rates from 0.38% up to 2.5%. Laboratory studies indicate rates varying from 1.1% to 21%. Again, these results are difficult to interpret – are they percentages of all cells, cells containing formulae, or unique formulae? (If a formula is copied down a row or column, it may count as many formula cells, but is only one unique formula). Taking a reasonably conservative line, suppose that on average 1% of unique formulae have errors – what does this mean for the overall spreadsheet?

Spreadsheets that I have seen recently have typically contained between 150 and 350 unique formulae. The probability of these containing errors would be between 78% and 97% on this basis. These rates are reasonably consistent with the overall error rates from the field audits. Probably about 85% to 90% of spreadsheets contain errors.

But my spreadsheets are better!

Many people will believe that these rates, though high, may indeed be accurate for the world at large, but that they (or their team) are much better than average. One actuary went so far as to say to me: 'As far as I am concerned, none of my spreadsheets has ever had a bug in it.'

Panko's study showed that people predicted about 18% of their spreadsheets had errors in them, while the true figure was 86%. This is a very general phenomenon. Experiments demonstrate that people generally think it likely they will live longer and

Table 1 Risk quantification methods

Source	Number of spreadsheets	Number with errors	Percentage with errors
Coopers & Lybrand, 1997	23	21	91%
KPMG, 1997	22	20	91%
Lukasic, 1998	2	2	100%
HM Customs and Excise, 2000	7	6	86%
Total	54	49	91%

healthier lives than their peers will, get better jobs, and earn more money. They also think they are less likely than their peers to develop drinking problems, be fired, be injured in the event of a natural disaster, or get divorced. About 80% of people believe they are better-than-average drivers.

People thus have misplaced confidence in their own spreadsheets. This confidence is probably reinforced by the fact that errors are rarely found, because so little testing is performed.

What about actuarial judgement?

I have heard it claimed that testing is not necessary because errors would be picked up through the exercise of actuarial judgement. If there is a mistake, the actuary will always spot that the numbers are wrong. I should certainly hope that major problems, such as the wrong order of magnitude, would be found, but what about results off by a much smaller amount? Surely if actuarial judgement were that good, we wouldn't need spreadsheets to do the calculations in the first place! A regular response to this is that small errors are unlikely to be significant anyway: this is simply not true. If you are pricing reinsurance contracts and regularly bid slightly more than your competitors, you may lose a lot of business. In the reverse situation you might win a lot of business that turns out to be rather less profitable than you had expected.

Actuarial judgement, while valuable, does not mean there won't be errors in your spreadsheets and is no replacement for proper testing.

Reviewing is not enough

We can see the risk is high. A large proportion of spreadsheets are likely to contain errors, leading to incorrect results. One of the reasons people don't test their spreadsheets is because they don't understand the magnitude of the risk.

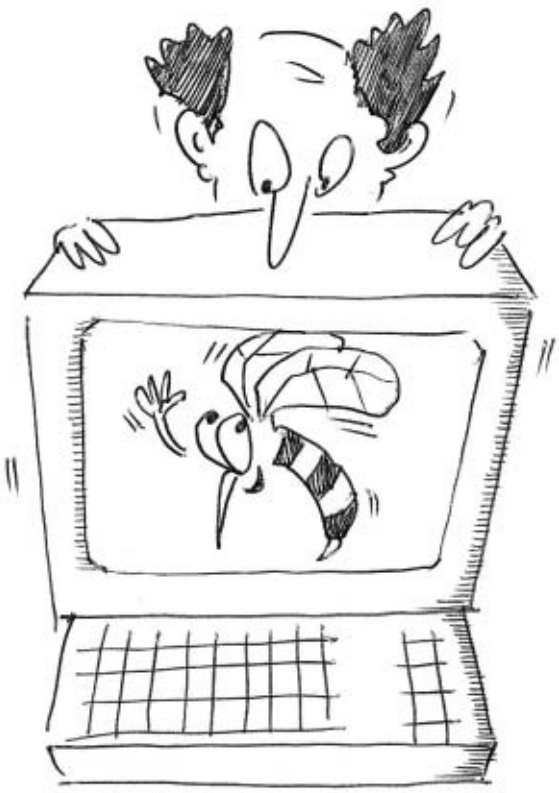
On the whole, though, people probably do review their spreadsheets (I may be over-optimistic here). Even better, in many cases spreadsheets are reviewed by somebody else. Why is testing necessary?

Testing is needed because reviewing, by its very nature, cannot pick up all the mistakes. As I mentioned above, it is widely accepted that this is so, and it's obvious if you think about it. The only way to make sure that a computer program works is to run it, and a spreadsheet is no different from any other computer program in this respect. You can review until you are blue in the face, but unless you perform the relevant calculations on a wide variety of inputs you can't be sure that the answers will be correct.

Systematic testing, with a comprehensive suite of test data, is simply the only way of ensuring the accuracy of your spreadsheets.

A cautionary tale

In March 2003 Provident Financial Group of Cincinnati announced a restatement of its results for the five financial years from 1997 to 2002. Between 1997 and



1999 Provident created nine pools of car leases. Part of the financial restatement was because the leases were treated off balance sheet, rather than on balance sheet as was later thought to be appropriate. But there was also a significant restatement of earnings, because there was a mistake in the model that calculated the debt amortisation for the leases. It appears that the analysts who built the model used for the first pool 'put in the wrong value, and they didn't accrue enough interest expense over the deal term. The first model that was put together had the problem, and that got carried through the other eight', according to the chief financial officer.

Apparently the error was found when Provident introduced a new financial model that was tested against the original, and found that the two models produced different results. They then went back and looked at the original model to see which one was correct. We don't know that these were spreadsheet models, but it's quite possible. And the lack of testing may have led to earned income being overstated by $70m over five years, as well as a class action lawsuit by disgruntled investors.

Can you afford not to test your spreadsheets?

In a 1% world it's more important than ever to control your risks. The risks of spreadsheet use are generally underestimated, but are reasonably easy to control. It cost Provident $70m not to test their financial models. Can you afford the risk? ❑

Louise Pryor is an independent consultant in user developed software and operational risk; she can be contacted through www.louisepryor.com