

Privileged and Confidential

Quantitative Approaches to Horizon Bugs

Summary

This note describes two quantitative approaches to the core issue of the Horizon trial, each based on engineering measurement. They are:

1. A 'micro' approach, which estimates quantitatively, for each error identified in Horizon, its possible net impact on the accounts of claimants' branches
2. A 'macro' approach, which estimates, for all possible errors in Horizon (discovered or undiscovered) their largest possible aggregate impact on the accounts of claimants' branches.

I believe that all these numbers can be estimated from evidence which can be made available; that estimating them is a necessary part of carrying out my expert duty; and that the resulting numbers will be small enough to refute the claim that bugs in Horizon account for a significant part of the claimants' losses.

This note is an addendum to our foundation report, further defining our approach to writing the full report. We would appreciate feedback on it from the legal team, PO, and Fujitsu.

Coyne's Strategy

I believe Jason Coyne has more or less revealed his strategy to me.

He has said that, since some claimants got into trouble for small amounts of money - as little as £2000 - any bug in Horizon, which could lead to an error in branch accounts, however small, will count. He is not assessing the size of the impact of any bug, just its logical possibility.

He envisages a joint expert statement which lists all possible bugs, regardless of the scale of their impact. He wishes to trap me into:

- a) **either** agreeing a long list of bugs, which goes on to be considered in a 'lead claimants' trial
- b) **or** claiming for some bugs that they could not possibly have led to any loss - a position he will be able to attack in his next report, for at least some bugs (leading to detailed debates which baffle the court).

Point (b) is a risk in spite of the triple filter analysis of bugs, which I expect will remove the majority of bugs- i.e. show they could not permanently affect branch accounts. The difficulty is that even with the triple filter approach, each bug on its own can be complex to analyse, reaching out into large parts of the software which we are not familiar with; and in the limited time available we may not be able to make all our analyses bullet-proof. A complex expert tit-for-tat may leave the court confused, and inclined to give the claimants the benefit of the doubt, in at least some cases.

I suspect Coyne intends to start with a list of, say, about 100 bugs in his first report - many of which he will not have told me about before he serves it - leaving me only a month to analyse, say, 50 bugs, before issuing my first report. This will not be enough time to do the thorough analysis needed for (b). So I will be left with a list that I will be asked to sign up to in an expert joint statement at some stage, not yet known.

If that is his strategy (and I can see few other viable strategies for him), I think the best way to counter it will include the micro and macro approaches outlined below. This proposal is based on my recent sampling of the 8000 KELs, and has implications for later trials.

'Micro' Quantitative Approach - Horizon Trial

From sampling a few claimants in the mediation, the final amount of discrepancy in their accounts ranges from small (say £2000) through a median figure of £20,000 - £30,000, to top-end figures of £80,000 or more.

From this, a conservative estimate of the mean discrepancy per claimant is of the order of £10,000 per claimant - leading to a total discrepancy across all claimants of £5M. There is undoubtedly a better estimate available, and the total can easily be computed. I shall use the figure of £5M for now. I shall call it the **Total Claimants' Discrepancy** (TCD).

The essence of the micro quantitative approach is to estimate - for any Horizon bug which might have affected branch accounts - what contribution it might have made to the TCD, on reasonable assumptions, and to express that amount as a percentage of the TCD. I expect these percentages to be very small, typically much less than 1% (see examples below).

The expected impact of any bug on the total discrepancy across all claimants will be called the **Expected Claimant Discrepancy** (ECD) for that bug.

A key conclusion of my first report will be a table of bugs - listing for each bug, *inter alia*, its ECD as a percentage of the TCD. I would expect the total of the ECD percentages will be much less than 10% - leading to the conclusion that all the bugs identified by Coyne are insufficient to account for even a small part of the claimed discrepancy.

What 'reasonable assumptions' can be made for the Horizon trial in computing an ECD? From our brief survey of KELs, the following ways of estimating ECD appear to be available:

- For some bugs, we can make a clear technical argument, using the triple filter framework, that its impact on branch accounts was zero, so its ECD is zero and it will have a zero entry in the table. This does not rule out an 'in the alternative' argument that, even if the bug had had some impact, its ECD is very small. The remainder of bugs are those we cannot fully eliminate by the triple filter in the time available to us.
- For some of the remaining bugs, FJ know the total money amounts involved. To compute the ECD of these bugs, we do not need to know the FAD codes of affected branches, or link them to claimants. All we need to do, to get an ECD, is to assume the bug was equally likely to affect all branches, claimants and others alike, and multiply the total amount identified by FJ by $(500/11,000) = 0.05$. We may occasionally need to vary this assumption, but not by far.

- For those remaining bugs where FJ do not know the amounts involved, we can proceed as follows:
 - From the KEL/PEAKs, find out the active period of the bug, between when it was introduced and when it was fixed or worked around - typically weeks or months
 - Find out the account codes or product codes which could be affected by the bug
 - Find out the (typically rare) circumstances in which the bug could occur (as nearly all bugs are rare edge cases), and express that as a probability per transaction of that type
 - From a sampling analysis of the 400 sets of transaction/audit records disclosed, (or from other factual data from FJ, e.g. from POL FS) estimate the typical annual volumes of money transacted in the identified account codes or product codes
 - To get the ECD as a monetary amount, multiply together:
 - the estimated total annual amounts of money passing through the account code
 - the probability of the circumstances of the bug occurring, when it was active
 - the length of the active period of the bug (expressed in years)
 - the factor (500/11,000) from assuming all branches are equally affected
 - To get the ECD as a percentage, divide it by the TCD and multiply by 100

For some bugs this will be an approximate calculation; but (by varying it as necessary) we should always be able to argue that it is the best approximation available within the constraints of the Horizon trial. I expect that the individual ECDs will be so small as to make their uncertainties immaterial. See the two examples below.

Micro Examples - Receipts/Payments Mismatch, and Suspense Account

For the receipts/payments mismatch bug, mentioned in Schedule 6 of the letter of July 2016, FJ identified the total amount involved as about £20,000.

In this case the calculation is simple. We only need assume that, in the absence of more detailed information (detailed information on affected branches actually exists for this bug, for possible use in the next trial - not in the Horizon trial), the bug was equally likely to affect claimants and non-claimants. The ECD is $£20,000 * 500/11,000 = £900$ (summed over all claimants). As a percentage of the TCD (guessed at £5M), this is 0.02%.

Similarly, for the suspense account bug, mentioned in the same schedule, FJ identified the total amount involved as around £10,000. This leads to an ECD of 0.01%

It is encouraging that two Horizon bugs, which have featured prominently in the case and are presumably prominent in the claimant's expert thinking, lead only to 0.03%, or one part in 3,000, of the TCD.

The claimants will need 1,000 more bugs like these, to get to 30% of the TCD. Or they will need to find 100 bugs, each of which has an ECD 10 times bigger than these bugs. This seems unlikely.

Work Required for Micro Analysis

In order to apply the quantitative micro approach to potentially a large number of bugs, we need as soon as possible to have:

- Access to Relativity, so we can rapidly investigate the special features of any KEL, to help us assess its probability of occurring, its potential impact on branch accounts, and the account codes involved
- Samples of transaction data and the tools FJ have developed to analyse them, so we can assess how useful the transaction data are in estimating the impact of the 'difficult' bugs, where a bug might affect branch accounts but its total impact is not known to FJ.

We need these things as soon as possible to assess and cost the approach.

We particularly need to find out about FJ's tools for analysing transaction data, to see whether they will meet our needs or whether we will need to develop some tools ourselves. While we expect that our analysis will be a fairly simple matter of filtering transaction records by account codes and time periods, and summing the monetary amounts, we urgently need to estimate the work involved in developing a working analysis process.

Micro Approach - Lead Claimants' Trial

In the later trial, which we assume will have a small number of identified lead claimants, we can estimate the impact of any bug on any claimant more precisely than the ECD:

- Time bracketing - could that claimant have suffered any loss, during the period he got into trouble? Is there any time overlap between the active period of the bug and the period of the claimant's difficulties?
- If the affected branches are known, was the claimant's branch affected?
- If not, examine audit records for claimant's branch during the overlap period, to compute the claimant's loss precisely.

So, for specific claimants, where audit data are available, we would usually expect to be able to reduce the ECD figure.

Some bugs could have ECDs so small that you could argue for them to be removed from consideration. So, the task facing the court could be simplified for this trial as well.

Macro Approach - Vigilant Postmasters

This section addresses two topics:

1. Quantifying the impact on branch accounts of any bugs in Horizon (the signal) in the presence of human errors in the branches (the noise)
2. The extent to which any software bug in Horizon, which affects branch accounts, would be detected and complained about by the more efficient and vigilant sub-postmasters

In measuring the signal, it would be too expensive to use data from all 11,000 branches. We would need to take a sampling approach. The most accurate way to measure the signal is to choose a sample so it has the least amount of noise compared to signal. This is standard engineering practice.

It is reasonable to assume that the level of the signal (errors in branch accounts caused by Horizon bugs) is approximately uniform across all branches, with only statistical fluctuations. So to measure the impact of Horizon bugs on branch accounts, one should look at the branches with the least noise (those with the fewest human errors; the most efficient and vigilant postmasters) rather than look at those with the highest noise (those with the highest rate of human errors; those who got into difficulties; the claimants)

How to choose branches with the lowest level of human error? Horizon is designed to correct for human errors, so that all human errors in the branches lead to one of two things:

- Correction transactions (though discrepancy accounts) authorised by the SPMR when rolling over to a new TP
- Transaction corrections, accepted by the SPMR

Therefore, we should define some metric (per SU, to allow for varying branch sizes) which combines these two factors; and in any given year, examine those branches (say, the top 50) which have the lowest level of that metric.

Then, if we assumed that those branches made no human errors at all (that their problems were **all** caused by bugs in Horizon; but that they accepted responsibility for all of them), the level of corrections in those branches gives an **upper bound** on the impact on branch accounts of bugs in Horizon. Having this upper bound is a first step; and it may be small enough to refute the main claim being put forward.

However, that is a very conservative upper bound, because those postmasters are the most vigilant and efficient ones. Whenever there is a discrepancy in their accounts, they are likely to investigate it thoroughly, and will only accept it if they are satisfied it arises from an error by them or their staff. If it arises from a Horizon bug, they will probably complain about it, and FJ are required to investigate it - leading to some combination of PEAKs and KELs.

We can expect that these postmasters run a tight ship. For detecting bugs in Horizon, they are the canaries in the mine.

How best to move from the conservative upper bound, to a fair estimate of the level of bug-induced errors in accounts? A first step might be to interview a small sample of the most efficient postmasters, asking them about their monthly discrepancies - are they satisfied that those arose from errors in the branch, or do they suspect other sources? What process of investigation did they follow?

Factual witnesses like this are probably out of scope for the Horizon trial, but interviewing them might yield clues as to other approaches, which can be evidenced from documents or data. For the later trial, just as the claimants are allowed a number of 'lead claimants' perhaps PO should be allowed the same number of 'lead non-claimants'.

In summary, if PO approaches the main issue of the Horizon trial as a problem in engineering measurement, and measures the signal from the sample with least noise, the answer may be low enough to refute the main claim. Common sense may be supported by statistics.

Tactics of the Quantitative Approach

At some stage, we would need to reveal this quantitative approach to (a) the court and (b) Coyne.

I believe there would be little difficulty in justifying the approach to the court, in terms of simplifying the task facing the court. In the micro approach for each bug, rather than diving into esoteric technical arguments as to whether the bug could affect branch accounts at all, the court needs only understand its range of quantitative impact - which is much more of a 'business' issue - and in most cases that range will be so small as to forestall the need for detailed analysis or statistics. The macro approach gives an independent check of the micro approach, which can be easily made.

In other words, the two approaches combined can be presented as a common sense approach. Rather than relying only on some abstruse expert analysis of architecture, code and error correction processes - which the court may have difficulty following - the approach is to isolate and analyse the data which most directly answer the key question. This is the approach most likely to yield a reliable answer within the costs and constraints of the Horizon trial.

The amount of factual data needed to support the micro approach is limited - perhaps limited to transaction data already disclosed, and some overall annual volumes of money by account code or product. Similarly the macro approach can be calculated using data from only a small sample of branches.

When the approach is revealed to the court, it will of course be evident to Coyne. I see no reason to tell him any earlier - except that I might reveal in broad terms that I am interested in the scale of impact of any bug. Or should I hold back even that? If we need to develop special analysis tools, is there any need to share them with Coyne?