Cashier Fraud: Data analysis in a context

by Jostein Lillestøl

Department of Finance and Management Science

The Norwegian School of Economics and Business Administration
Bergen, Norway

Summary

A true story is told about the role of a statistical expert witness in a cashier fraud case. It illustrates Shewhart's first principle of understanding data: " Data have no meaning apart from their context". It may be used in the classroom as is, or the context may be changed as indicated.

KEY WORDS: Fraud, evidence, variation, teaching

Introduction

This paper deals with cashier fraud at a supermarket, involving so-called "return money" when correcting errors. Being called upon as a statistical expert witness in a court case, the question was to provide a lower limit on pocketed amount, taking into account natural variation on return money within and between cashiers over time. Data on daily returns for the defendant and her colleagues were available over a period of a year and a half. The analysis could proceed under different assumptions, giving rise to more or less conservative conclusions. The mode of analysis was very much dependent on the context. It also triggered interesting questions on the use of statistics as evidence.

The crime scene

The supermarket in question has 16 cashiers, about 8-9 at work at a time. Each cashier has her/ his own code that is entered in the computerised cash register before operation. The cash registers has two rolls, one producing itemised receipts to the customer, the other recording the same amounts without item specification to be kept by the store (called gossip roll). Typically such rolls are replaced (more than) once during a workday. If an erroneous amount is entered into the register or merchandise is returned, it shall be detracted from the register again, and then recorded in writing on a return/correction sheet. The wrong entry and its correction will then appear on both the customer receipt and the gossip roll. At the end of the day the cash in hand and the return/correction sheet, was brought to the back office to be balanced. There was no day-to-day inspection of gossip rolls.

The crime

One of the cashiers (A) seized the opportunity to deliberately enter erroneous amounts. These were not corrected in full on the rolls, but recorded as if on the return/correction sheet. The difference was then pocketed. According to the routines the cash at hand balanced the records at the end of the day. To reduce the risk of gossip rolls telling the truth, the parts in question was occasionally scissored off. The discovery of fraud by inspecting these rolls was then dependent on discovery of time gaps between successive rolls.

The discovery

In May 1993 the supervisor surveyed the sales turnover and return money for each of the cashiers during the first 4 months of the year. She discovered that cashier A had fairly frequent returns and higher return amounts than other cashiers. The return/correction sheets were then compared with the gossip rolls, and mismatches were discovered. The complete records from the previous year added to the suspicion. By after hour inspection of the garbage cans the management found a piece of a gossip roll of cashier A. Nevertheless her cash balance was apparently correct, which was taken as an evidence of withdrawal of money that day. A complaint was then filed to the police.

The concession

The above findings together with statistics from the computerized cash register was basis for the examination by the police. The cashier quickly admitted having embezzled varying amounts, ranging from NOK 1000 up to, may be, NOK 10 000 a time on average once a week for a little more than a year and a half (1 £ = 10 NOK). Concerning the total amount she was confronted with a calculation covering the period 03.10.91 to 07.05.93 in which her return money amounted to about NOK 400 000 compared to an average of NOK 50 000 for all cashiers, thus leaving NOK 350 000 for her to account for. Although the accused could not imagine having embezzled this large an amount, the police obtained her concession based on «undisputable facts».

The statistical prosecutor

The prosecutor to be quickly realised a possible flaw in using this calculation as basis for the accusation: The accused may have lower abilities than average at the outset, and should not be punished for that. In a criminal case «every penny» embezzled must be proved, and the average argument would be a gift to the defence. After preliminary consultation of statistical expertise taking the variation of return money between cashiers into account as well, the accusation was reduced from about NOK 350 000 to NOK 300 000. A professor of statistics was wanted as an expert witness, since statistical arguments of this kind had not previously been brought to the court. The commission was given: «To analyse the information made available and establish how much the return money of the accused deviates from natural variation among the cashiers in the store. State the assumptions of the analysis, and how sensitive it is to realistic changes of the assumptions».

The data

The available data included individual return amounts, number of returns, sales turnover among others, for every cashier and every work day in the period 03.10.91 to 07.05.93. The most important condensed numbers were:

Cashier	Return NOK	No. of returns	Averaging
A	400 611	804	498.27
B	18 574	390	47.63
C	70 730	774	91.38
D	88 473	938	94.32
E	49 201	769	63.98
F	55 142	873	63.16
G	47 919	1 306	36.69
H	34 915	758	46.06
I	6 030	115	52.43
J	69 758	1 574	44.32
K	5 604	141	39.74
L	52 882	387	136.65
M	3 312	72	46.00
N	27 613	713	38.73
O	109 812	1 064	103.21
P	41 892	720	58.18

Of course some cashiers had worked more than others and data on sales turnover and numbers of items entered were available for analysis as well.

The report

The available data supports the following:

Average consideration:

Defendants total returns	400 611
«Expected» return amount	51 592
Deviation (unexplained)	349 019

The expected return is computed from the defendants 804 returns, under the assumption that the defendant at the outset is on the average of the 15 cashiers in the control group. Their average is NOK 64.17 per return compared to NOK 498.27 of the defendent.

Conclusion

The data has made plausible an embezzlement of about NOK 350 000.

Favour consideration based on normal individual variations

It may be the case that the abilities of the defendant was different from the others at the outset. An estimate of most likely individual variations in return amounts is +/- 23 400, computed from the standard deviations of amount per return among cashiers of NOK 29.19 in the control group scaled up to the 804 returns of the defendent.
To give favour to the defendant according to a normal distribution of abilities, amounts to reducing the unexpected deviation above by *k* * 23 400, where *k* is dependent on how much favour to be given. The following table shows some cases:

k	Favour	From bottom	Reduction	Deviation (reduced)
0	No	50 %	0	350 000
1	Moderate	16 %	23 400	326 600
2	Large	2.5 %	46 800	303 200
3	Extreme	0.13 %	70 200	279 800

Here the right hand column is an estimate of embezzled amount for various degree of favour. Extreme favour (only about 1 of 1000 cashiers comparable to the control group have abilities below this level) corresponds to an estimate of about NOK 280 000.

Alternative favour considerations could be made by taking the extremes in the control group itself as benchmark, but this gives less than extreme favour according to the normal consideration:

i) Computation based on cashier (L) in the low end of control group (who also is somewhat atypical) gives 400 000 - 804 x 136.65 = 290 000.

ii) Computation based on return amount in per cent of own sales turnover (NOK 8 005 000) gives:

Normal return in control group:	8 005 000 x 0.54 % = 43 000
Extreme return (cashier L):	8 005 000 x 1 % = 80 000
Deviation:	400 000 - 80 000 = 320 000

Conclusion
Different favour considerations gives a deviation from natural variation ranging from NOK 280 000 to NOK 350 000 (no favour). How much favour to be given is up to the court to decide, also taking the considerations below into account.

The abilities of the defendant
Date for successive time periods may document the normal abilities of the defendant, and also indicate when the embezzlement started and how it developed.

In the table below is information about number of returns and average return amounts in 7 successive time periods starting on the time she was employed as cashier (the first six having about 40 workdays, the last having 30 workdays).

	1991	1992				1993	
No. of return	155	118	106	102	109	126	88
Amount	235	299	299	616	641	819	693

Notice that the number of returns decreased after her first period of employment, indicating increased abilities. Nevertheless the average amounts show an increased tendency over time, with a dramatic jump in the middle of year 1992. Notice however that the average return amount is already higher in the first period than for the other cashiers. This could indicate that she is in some sense different from the others at the outset. This point is however weakened if the data from 1991 is split into two periods: The 10 first workdays had an average of NOK 40 (55 returns) increasing to NOK 343 (100 return) in the next 30 workdays. The difference is statistically significant, that is, not likely due to chance alone. The return amount the first two weeks were in line with the other cashiers. This may raise the suspicion that the embezzlement started earlier than conceded. However some caution must be taken: The special work situation of a newly employed may cause special behaviour not reflecting normal working capacity.

I see no other support for assuming that the defendant may have had abilities different from the others at the outset. For instance the number of returns compared to own sales turnover and number of items entered is in line with the others. Five cashiers in the control group have in fact more returns compared to the number of items entered.

Conclusion
There is no basis for assuming that the defendant at the outset had a capacity different from the other cashiers. This means that there is no compelling argument for giving the defendant favour beyond the observed individual variation among the cashiers in general expressed in section 7.2.

The Court Proceedings

An experienced judge administered the court, having two laymen.

The size of fraud settled on by the court would be crucial to the defendant, since the amount qualifying for more severe punishment is somewhere in the range of NOK 200 000 - 300 000.

Prior to the proceedings the prosecutor was not sure how the court would react to statistical arguments, they could just as well become confused by them. Consequently some effort was also spent on the high returns which was individually checked.

Testimony was given by the store owner and by the store manager, who also provided compelling bar charts of some key statistics (of skyscrapers among bungalows, according to a local newspaper). The statistician then gave a brief summary of his report.

Based on the evidence and testimony the prosecutor advocated a fraud of NOK 300 000 as proved beyond reasonable doubt. The defender did not dispute the statistics at all. However the defendant confirmed a question by defence that she occasionally entered and corrected some high amounts as to cover up, by making such entries more likely. The prosecutor said afterwards that he at this point feared that this possibility could have destroyed his case, despite the compelling statistical evidence. The store manager however denied that correct high returns existed at all and this closed the argument.

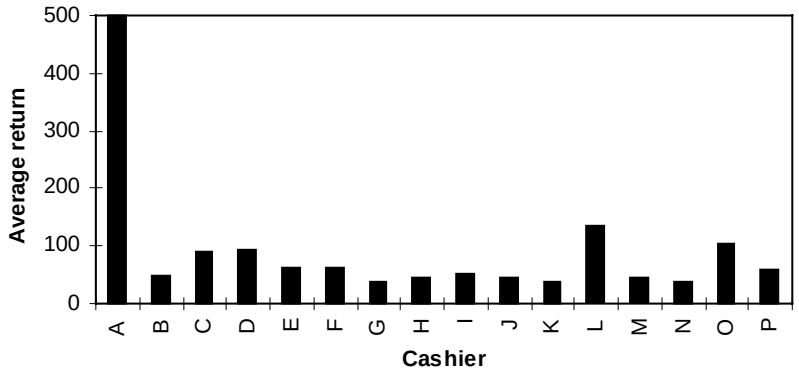
Although we have no definite knowledge, we have reasons to believe that the judges understood and valued the statistical arguments, despite the fact that the defendant was given the added benefit of doubt by settling on a verdict of NOK 250 000.

Afterthoughts

The story told is a true story about a context and a mission for a statistical expert witness, and a true story about how the mission was carried out. The mission provided an opportunity to reflect on the use of statistics as evidence, and on the importance of understanding data in a context. With the available data a more sophisticated analysis was possible, for instance by looking at the data in a time series context. However, by introducing modes of analysis that are harder to explain to the court, you may increase the risk that the statistical arguments are dismissed altogether.

The store supervisor, the police investigator, the prosecutor and the defence lawyer could all make use of the same data in their context, but pose different questions and use different modes of analysis, depending on the available supplementary information.

For the store supervisor the graphics below would be sufficient to raise suspicion and go ahead. The statistical expert may regard this as too persuasive to be used in court.



An interesting and important question is whether and how things will be different, prior to court proceedings and in court, if the accused had admitted nothing. This has to do with the limits of statistics as proof. One could argue that occurrence of an unlikely event proves nothing. Take the lottery analogue: Someone will always win the top price. It does

not make sense for the lucky one or the unlucky ones to compute a probability afterwards to advocate that the outcome should not happen.

In the given context the case is about variation and not about confidence intervals and not about outlier detection either. In a presentation to a statistical audience the author was met by several comments linked to these misconceptions. It is maybe a symptom of our profession that we are likely to phrase statistical problems in our inference terms, whether it is relevant or not. It is my sincere opinion that we should talk a lot more about variation and less about formal inference.

The context and the mission may be given as a student exercise as it is. Afterwards the complete story can be revealed for discussion and critique. The case could alternatively be discussed from the point of view of the store supervisor, the police investigator, the prosecutor or the defence lawyer. You could even imagine role-playing. For the role of store supervisor you set the crime scene and reveal the suspicious circumstances and data, and then ask the student how to proceed.