When Numbers Lie: Good Bad Examples

Susanne Ditlevsen ·
Andreas Winther Jessen ·
Rolf Poulsen

February 9, 2021

Abstract Numbers don't lie, says the old adage. But if tortured sufficiently, they will say anything. We give examples from finance, public administration – and a store front window.

Keywords Computational errors · public tender · interest rate and credit risk

1 A very short introduction

We give a number (small, but larger than it shoukd be) of examples of innumeracy (i.e. a lack of ability to understand and use numbers in calculations) and other bad principles in finance and surrounding areas. The examples are not very technical, but they are very important. Moreover, they are (a) real, we have made nothing up, (b) freighteningly recent.

2 Starter for ten: Percentages, distributions, and negative numbers

The three concepts mentioned in the section title are usual suspects in cases of quantitative confusion. From a long list, we pick three examples of recent lived experiences.

Percentages. Expressing effects in relative terms can be very useful for intuition, but we must be careful when competing scales are in use and they

Ditlevsen and Poulsen (corresponding author, `rolf@math.ku,dk`)
Department of Mathematical Sciences, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark

Jessen
Copenhagen Simulations, Colbjørnsensgade 25, DK-1652 Copenhagen V, Denmark

Fig. 1 Decrease in human body temperature expressed as a percentage. Not useful without mentioning which temperature scale is used.

Fig. 2 Illustration of an average rate (*gennemsnitsrente* in Danish) investment product. The average should be on the x-axis, not on the y-axis.

do not agree where to put zero. Over the last 150 years the average human body temperature has fallen from 37.0 degrees Celsius to 36.4 degree Celsius. One hypothesis – but not the only one – is that is modern medicine, antibiotics in particular, means that fewer people are in a state of permanent infection. The tweet shown in Figure 1 expresses this as a 1.6% (=100*0.6/37.0) drop. Howver, someone using the Fahrenheit scale (and a base temperature of 98.6 degrees F) would see the relative change of 1.6% as an absolute drop of 1.6 derees Fahrenheit – which is a about 0.9 degrees Celsius, not the 0.6 degrees Celsius from the original fact. Someone using the Kelvin scale would see a drop of almost 5 degrees Celsius. The problem here is that the zero on the scales do not represent the same temperature, so we can't use the percentage information without knowing the scale.

Distributions. There are good reasons why a pension saver will want a plan that smoothens the volatile rates of return in financial markets; an average (*gennemsnit* in Danish) rate product. Many pension companies offer such products. One major Danish pension company (more than 15 billion euro under management) chose to illustrate their version on their store front window in the way shown in Figure 2. The astute reader will note that: "No, the average goes on the other axis; the line should be vertical". (At least we think it should, but strictly speaking it is hard to tell without units/labels on the axes.)

Negative numbers. Interest rates used to be positive, sometimes very much so. That is not necessarily the case anymore. For instance, in the Danish mortgage market floating rate loans have been in negative interest rate territory for a handful of years and in 2020 you could issue a 0-rate 20 year callable,

Du får her en beregning af lån i ejendommen█████████████████████████
Beregningen er lavet på F-kort, 30 år, med afdragsfrihed. Lånet er specificeret efterfølgende.

Overblik   (Hovedstol 1.500.000)

Ydelse pr. år 1. år før skat	8.984
Ydelse pr. år 1. år efter skat	8.541
Heraf afdrag pr. år 1. år	7.082

Gennemsnitlige årlige omkostninger i procent udgør før skat 0,18% og efter skat 0,16%.

Fig. 3 Break-down of the first year payments on an interest-only mortgage loan. The first and second lines of the body of the table show, respectively, the before and after tax (*skat*) first year payment (*ydelse*) on a DKK 1.5M loan. The third line shows the amout that goes to reduction of the principal (and should be zero if the loan were truely interest-only).

fixed rate bond at very close to par. Negative interest rates have some counterintuitive theoretical effects and give practical challenges; some cash-flows (should) "go the other way". Figure 3 gives a concrete example. It is a bank statement showing the first year payments on a 30-year floating-rate, interest-only loan, meaning no repayment of principal for the first 10 years. However, the statement shows a first year (after-tax) payment of 8,541 Danish kroner of which 7,082 (83%) goes to reducing the outstanding principal. That ought to be zero on an interest-only loan. The reason is: The market conditions are such that interest rate on the loan is negative (about -0.5%), so the borrower should recieve money from the lender. However, that does not happen, instead the interest rate payment that the borrower should have recieved is used to reduce the principal (*afdrage på lånet* in Danish). There are sound reasons for this (one is that it can all be done administratively by the bank; the borrower does not have to transfer money), but without any explanation, the statement makes for quite confusing reading.

3 Pieces of eight: How not to do a public tender

The examples in the previous section were benign: Bad communication, sure, but nobody got hurt. Not so with our next case.

In the autumn of 2019 the Danish government solicited bids for a nationwide digital radio channel; the DAB-offering for short. The winner of the offering would receive public funding of up to 280 million Danish kroner (DKKM; around 38 million euro) over four years. Applications would be assessed and the winner chosen by the Radio & TV Board (the Board in the following), an independent 7-person board under the Ministry of Culture. The Board members' backgrounds were in law or economics - none in publishing, and seemingly none with a decent level of numeracy. Applications were (supposed to be) scored in three main categories with weighting and subcategories, where

each subcategory scored an integer between 0 and 8, as published in the official description of the public tender.[1]

– Operating plan (administrative stuff) with overall weight 35% and 3 sub-categories.
– Budget (amount in DKK asked for in the application) with overall weight 25% and no subcategories.
– Program plan (the creative plan) with overall weight 40% and 11 sub-categories, with one subcategory given special weight relative to the other subcategories.

By the deadline on September 22, 2019 there were three applicants; Radio24syv, LOUD, and dk4. On October 22, the Board announced its decision in a report[2] whose main conclusion is in Table 1.

Table 1 Summary of bid assessments in a Danish public tender round for a nationwide digital radio channel.

	Radio24syv	LOUD	dk4
Operating plan	7	6	3
Weighted point - 35%	2.45	2.10	1.05
Budget	1	7	8
Weighted point - 25%	0.25	1.75	2.00
Program plan	6	7	4
Weighted point - 40%	2.40	2.80	1.60
Total score	**5.10**	**6.65**	**4.65**

From the last row it follows that LOUD was the winner - and seemingly a clear one at that. This came as a surprise to many. Radio24syv was generally regarded as the favourite, having run very successfully a similar channel on FM radio for the previous eight years. However, looking at the numbers and calculations in the Board's report, problems started emerging. A cocktail of numeric errors and dubious algorithms had systematically been implemented in the evaluation.

3.1 When to round off, when to average, and how not to weigh

For each applicant and within each category, the points from the subcategories were added up, divided by the number of subcategories, rounded to the nearest integer, and multiplied by the category's weight factor. That number was the final points obtained in the given category, and this was added to the points obtained in the other two categories, giving the total score in Table 1. However,

---

[1] The official document outlining the rules of DAB-offering is here. In Danish only – as all official documents in the case.

[2] The report with the Board's decision here.

this violates the most basic rule of numerical analysis: *You do not round off in intermediate calculations.*[3] If – as here – you do round off, you lose significant accuracy for the final result. In the category Program plan, the raw numbers were that Radio24syv scored 6.3 points, LOUD got 6.6 points and dk4 got 4 points. So there is only a difference of 0.3 points between Radio24syv and LOUD. After weighting with 40 %, this becomes a difference of 0.12 points. But that's not how the Board calculated it. They rounded to the nearest whole number before applying the weights - therefore Radio24syv got 6 points and LOUD 7 points. Now there is suddenly a difference of one whole point – which after weighting becomes 0.4 points. It is this difference that is used in the calculation of the final score. So there ends up being a *larger* difference in absolute numbers between two applicants after down-weighting the result by 40 % than in the original numbers! Note that if all three applicants had scored either 0.2 points more or fewer, corresponding to a translation of the scale - then Radio24syv and LOUD would have received the same score - namely either 7 (if they had all received 0.2 points more) or 6 points (if they had all received 0.2 points less). Thus, the *relative* scores between applicants depend on the chosen scale. In this case the rounding happens to exacerbate the difference between LOUD and Radio24syv. More generally it is (under reasonable assumptions and from a simulation experiment) similar to adding uniform random numbers between -5 and 5 to the subcategory sum totals. In short, while intermediary calculations are rounded to one digit ($n \in \{0, 1, \ldots, 8\}$), the final results are provided with (what is here) three significant digits ($n.xy$) – at least two of which are then noting but random noise.

The rounding issue was not the only case of flawed numerics: One of the Program plan subcategories (Culture) was to be given double weight relative to the other 10 subcategories. The Board did this by multiplying the applicants' point scores in this subcategory by two when calculating the subcategory sum totals. When the averages were calculated, the sum totals were divided by 11, the number of the subcategories. However, because the Culture subcategory effectively enters twice, the sum totals should be divided by 12. It might be less obvious than with the rounding why this is wrong, but consider this example:[4] You take two exams, both scored from 0 to 100. One has double weight. You score 80 on both. Of course, your weighted average score should be 80. By the Board's method, your weighted average score would be (2*80+80)/2=120 out of 100.

3.2 The budget evaluation – a tale of many errors

The errors described in the previous subsection are annoying to us and many others, including readers of this journal we hope, but they are not malicious;

---

[3] Unless you absolutely have to – and in that case the problem is most likely as completely different kettle of fish.

[4] We thank Carl Winsløw for providing the example of the exam scores.

it's equivalent to adding random noise. We are not sure that be said for the problems we now describe.

How to give more weight to the budget category than intended Table 2 takes a closer look at the Budget category. Radio24syv applied for 100% of the

Table 2 Summary of applied budgets and the points assigned to each applicant in a Danish public tender round for a nationwide digital radio channel. The total score is the numbers appearing in the budget category of Table 1.

	Radio24syv	LOUD	dk4
Million DKK applied for	280.00	260.82	256.25
Pts., interpolation formula	1.03	6.66	8.00
Pts., rounded	1	7	8
Total score, weighted 25%	**0.25**	**1.75**	**2**

maximal possible funding and got almost the minimal score, 1 out of 8. LOUD and dk4 applied for, respectively, 93% and 92% of the maximal funding and received scores of, respectively, 7 and the maximal 8. Intuitively, that seems like a large swing in points based on differences in amounts that are small compared to the total budgets, less than 9%.

This is a deliberate part of the construction of the DAB-offering. The lowest bid will always obtain a maximum score of 8, independently of how small the difference is to the highest bid. The highest bid will get a score that depends on the minimal realistic bid, which in this case was decided to be the lowest bid rounded down to the nearest 10 DKKM, in this way forcing the lowest point to be close to zero. We will come back to that. Constructing an evaluation that forces this category to use the entire scale from a value close to 0 up to the maximum of 8, whereas the other categories only use the central part of the scales will effectively give the budget category a weight that is much larger than the announced 25%. Let's see how this works. A bid of b is awarded $p(b)$ points, where $p(b)$ is given by the piecewise linear interpolation function

$$p(b) = \left(8 - \frac{8 - 0}{s} \frac{b - b_{min}}{b_{min}} \right)$$

where 8 and 0 represent the upper and lower point limits, b_{min} is the minimal budget applied for across all applicants, and the so-called slope coefficient s is determined as the simple average of two terms,

$$s_1 = \frac{\text{maximal possible funding (280 DKKM) - minimal realistic bid (250 DKKM)}}{\text{minimal realistic bid (250 DKKM)}}$$

$$s_2 = \frac{\text{maximal bid (280 DKKM) - minimal bid } (b_{min})}{\text{minimal bid } (b_{min})}$$

Figure 4 shows the interpolation formula, calculated with the minimal realistic bid used by the board, as well as using the values of 50% and 80% of

the maximal possible funding, respectively. The figure shows how much the differences in scores depend on this value. Even if changing this value does not change the order of the scores obtained by the applicants, it will hugely influence the *differences* in scores between applicants, which will be decisive when weighted and summed with the other categories. Choosing a minimal realistic bid that is approximately equal to the minimal bid will induce a large range between scores even if the bids are only differing by small amounts, and thus, the budget category will most likely be the decisive factor even if weighted low. Notice in Table 1 how the supposedly most important category weighted with 40%, the Program plan, only varies between 4 and 7, and thus, after weighting, only varies over 1.2 points, whereas the Budget category weighted with only 25% provides differences in final points of 1.75.
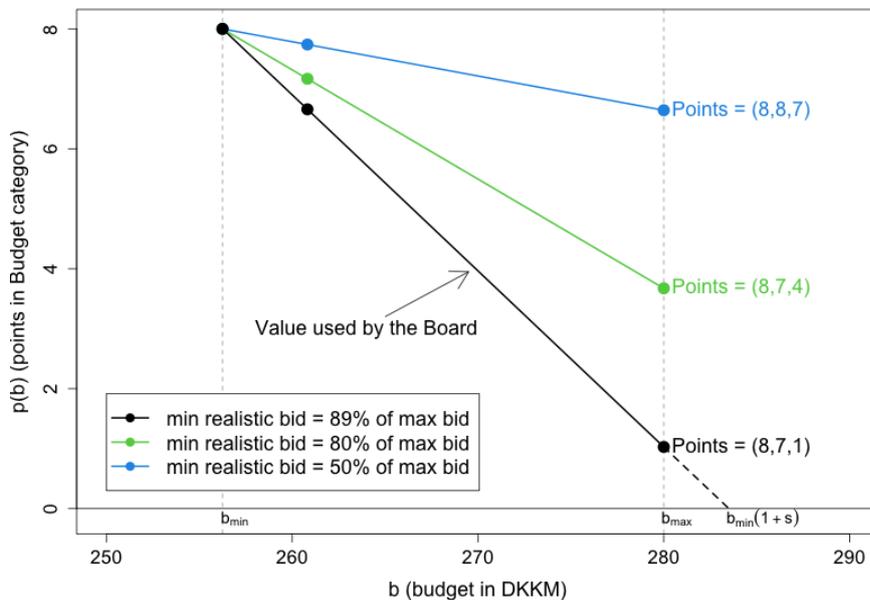
Fig. 4 The interpolation function used for awarding points in the Budget category. The slope depends upon the chosen value for the minimal realistic bid. The black curve was the choice of the Board, made after having seen the bids. The blue and the green curves are with values of 50% and 80% of the maximal possible funding (b_{max} = DKKM 280), respectively. The dots indicate the bids of the three applicants, with the lowest bid by dk4, and the largest bid by Radio24syv. The maximum score difference between the applicants change from 7 to 1 points for the given values of the minimal realistic bid.

How to manipulate the scores by deciding the minimal realistic bid after seeing the applications The Board did not specify in advance the minimal realistic

bid. This was considered part of the competition. However, the resulting scores depend heavily on this value, in fact, the scores can be manipulated by varying this value, see Figure 4. Nevertheless, the Board did not ensure that their minimal realistic bid, while not revealed in advance to the applicants, could be verified independently afterwards by a third party. In fact, the Board set the minimal realistic bid size (at DKKM 250) at its meeting on October 22 after having seen and scored all the applications, which they stated openly.[5] They further argued that the ordering of the scores would be the same regardless of the value of the minimal realistic bid, completely ignoring that the differences between scores matter when weighted together with the other categories. *It is simply not an acceptable procedure.* Whether the minimal realistic bid was decided upon to manipulate the outcome, or whether it was simply due to innumeracy in the Board is unknown.

How the difference between two applicants depend on the bid of a third applicant. There is another subtle problem build into the construction of the budget evaluation through the interpolation formula. Since dk4 has scored very low on all parameters other than the economy, they will never be able to win the bid. The interest therefore gathers around Radio24syv and LOUD, which are in a close race. So what we really want to compare is Radio24syv and LOUD. They ended up with a difference of 1.75 points after weighting in the budget category. Let us assume that the minimal realistic bid is fixed at 250 DKKM. How would the point allocation have been for Radio24syv and LOUD if dk4 had bid 6% lower, namely 240 DKKM? Since the smallest bid is included in the calculation of the slope, the point allocation after weighting would now be 0 points for Radio24syv, 0.8 points for LOUD and of course still 2 points for dk4 (it is still the smallest bid). Now the difference between Radio24syv and LOUD is only 0.8 points in the economy, against 1.75 before! Let's just understand this: The relative rating of the two relevant bids depends on what a third, not relevant bid is!

Readers wanting hand-on experience with the various perverse effects of the full algorithm for converting (all) points to (one) winner can use this spreadsheet.

4 Market values without markets: Interest rate and credit risk

During and shortly after the financial crisis in 2007-9 many small-to-medium companies (500-1000 in Denmark alone) changed their traditional floating-rate bank loans to fixed rate bank loans. Technically, this was done with interest rate swaps, but here we can focus on the net outcome: The company has a (typically long term; 20-30 years) fixed rate loan with the bank as counterparty. By swapping to a fixed rate loan, the company has removed all financial risk,

---

[5] See https://tinyurl.com/qt56fja, answer 68 specifically.

right? Wrong! The company is exposed to two types of financial risk, which we will discuss in term: Interest rate risk and credit risk.

Interest rate risk. Having a fixed rate loan means that you know how much you pay every (say) year. However, the value of your debt (i.e. the sum of the discounted values of all your future payments) will fluctuate with the market interest rate. If the interest rate goes down, the value of your debt goes up. This is "value" in a quite real and unambigous sense; it is what it will cost you today to get out of all your future obligations. As a numerical example, suppose at time 0 you issue debt in the form of a 30-year bullet bond with a coupon rate of 5% and principal DKKM 25 (a typical value). Assume further that the market interest rate is also 5% (i.e. the yield curve is flat at 0.05). The bullet bond would trade at par, you would get DKKM 25 today to run your company, and the value of your debt would be DKKM 25. Let's say that after 10 years the market conditions have changed; the yield curve is now flat at 2%. The value of your debt is

$$\text{DKKM } 25 \left(\frac{1}{1.02^{20}} + \sum_{t=1}^{20} \frac{0.05}{1.02^t} \right) = \text{DKKM } 37.26,$$

an increase of about 50% or in monetary terms a loss of more than DKKM 12. This seems counterintutive; you try to remove risk and end up with a loss of 12M because of market movements.[6] In many cases companies said "We were not informed of that risk by our bank" and have taken the bank to court. To briefly summarize (largely based on anecdotal evidence and personal experience, as many cases are settled without a public verdict), the advice of banks mostly holds up in court, but only just. In the Danish mortgage market this downward interest rate risk (to the borrower, i.e. the individual house owner) is alliviated by the bonds issued being callable; the borrower can get out of her future oligations at any time by prepaying the remaing principal. The risk when foregoing callability (which can be done by issuing bonds with a very low coupon rate) has been known since the late 80'ies and is reflected in the rules of thumb used by mortgage advisors, see [1].

Credit risk. Let's look again at what happens at time 0 in our numerical example. In the real world a bank would not pay out DKKM 25 to the company for a loan that has a 5% coupon rate. It would demand a higher interest rate, for instance 7%, reflecting the default risk of the company and the bank's (running) costs, as well as its profit margin. Even though the 2% spread might be a perfectly fair value, the bank will typically just say "you must pay 7%", not specify something like "of which 5% are market and 2% are for risk and costs". Now for the situation after 10 years. Standard practice of Danish banks

---

[6] One can prove that for a floating rate loan, the value of the debt (in an ideal world) is always equal to the remaining principal.

has been to report the value of the debt as

$$\text{DKKM } 25 \left(\frac{1}{1.02^{20}} + \sum_{t=1}^{20} \frac{0.07}{1.02^t} \right) = \text{DKKM } 44.45, \tag{1}$$

i.e. almost DKKM 20 over the initial payout; a large loss to the company. But is this a reasonable practice? The company would argue that it is not fair to include the (full) spread in the calculation (having 0.07 in the numerators); if the company repays now, there is no future credit risk for the bank, the running costs also disappear, and the bank can invest the money elsewhere to cover its profit margin.[7] Thus, the fair value of the debt should be closer – if not equal – to the DKKM 37.26 from earlier. To this, banks have effectively replied "fair-schmair – 7% was the deal you agreed to". However, there is the contractual stipulation that the company can always terminate the swap agreement at market value. But as there is no directly observable market price for this very specific swap between the company and the bank, it has to be estimated in some way. Banks report the number calculated in equation (1) and refer to it as 'market value'. However, that goes against the internationally accepted accounting principle that the market value of something is the amount that the bank could sell it for to another (similar) bank. And that would definitely not be DKKM 44.45 because the buying bank would demand compensation for credit risk etc. So what should the market value be? Well, that is where it gets tricky. A unique answer cannot be given without further modelling assumptions, both about the company, the bank, and the interest rate market as well as additional calculations. This is what is commonly known as xVA; VA for value adjustment, x to cover different aspects (C for credit, F for funding, …). It has been a very active field for several years, see for instance [2], and xVA calculations in some form have been implemented in any bank of repute.

So, in summary, it *is* sensible to report the number from equation (1) because it reflects changes in the interest rate market and is directly verifiable, but it *should not* be used on stand-alone basis, and it certainly isn't the market value of the debt.

5 Post scriptum: Light in the gloom

A natural question is now: Has anything come of the various criticisms in this paper, or are we screaming into a void? Looking into that allows us to end on a positive note.

The DAB-offering The decision caused quite a stir in the Danish media, and the points raised above (and more, related to the content of the evalua-tions, but that is another story) were presented to the Board — which did not

---

[7] To compare, again, to the mortgage market, borrowers that prepay there do not have to compensate the mortgage bank for future non-existing costs.

respond in a particularly constructive manner. One of their answers was:[8] "We acknowledge that the calculations could have been more mathematically correct, [...] but it wouldn't have mattered for the final result". You can't argue in this marginal adjust-things-one-at-the-time way; you must fix everything that is wrong and then see what the combined effect is. However, as time went by, the flaws in the tender process gained 'political traction'. Nobody would (or could; the legal details are subtle) change the decision in the DAB-offering case, but in late 2020 a number of changes (different procedures, new board members) were implemented in response (implicitly, if not explicitly) to the criticisms raised.

Market values of swaps. From the summer of 2020 this story was also covered extensively in the Danish media – albeit different parts of the media than found interest in the DAB-offering. Despite the reluctance of banks, in December 2020 the Ministry of Industry, Business and Financial Affairs stipulated that reported debt values should include xVA-adjustments, leaving the specific details to the parties involved. In January 2021 the Danish Association of Chartered Accountants introduced it as official practice, again with the finer details to be determined.

References

1. Rasmussen, Kourosh Marjani, Claus A. Madsen & Rolf Poulsen (2014), "Can Home-Owners Benefit from Stochastic Programming Models? Study of Mortgage Choice in Denmark", *Computational Management Science*, vol. 11(1-2), pp. 5-23.
2. Gregory, Jon (2020), *The xVA Challenge: Counterparty Risk, Funding, Collateral, Capital and Initial Margin*, Wiley.

---

[8] See https://tinyurl.com/qt56fja, answer 61 specifically.