

CONTRIBUTED PAPER

The Global Financial Crisis – risk transfer, insurance layers, and (lack of?) reinsurance culture

Michael Fackler

Actuary (DAV), Munich, Germany
Email: michael_fackler@web.de

Abstract

The Global Financial Crisis of 2007–2008 has elicited various debates, ranging from ethics over the stability of the banking system to subtle technical issues regarding the Gaussian and other copulas. We want to look at the crisis from a particular perspective. Credit derivatives have much in common with treaty reinsurance, including risk transfer via pooling and layering, scarce data, skewed distributions, and a limited number of specialised players in the market. This leads to a special mixture of mathematical/statistical and behavioural challenges. Reinsurers have been struggling to cope with these, not always successfully, but they have learned some lessons over the course of more than one century in business. This has led to certain rules being adopted by the reinsurance market and to a certain mindset being adopted by the individuals working in the industry. Some cultures established in the reinsurance world could possibly inspire markets like the credit derivatives market, but the subtle differences between the two worlds matter. We will see that traditional reinsurance has built-in incentives for (some) fairness, while securitisation can foster opportunism.

Keywords: Financial Crisis; Risk Transfer Chain; LMX Spiral; Model Uncertainty; Incentives

1. Introduction

1.1. Motivation

The Global Financial Crisis of 2007–2008 (GFC) has generated, beyond the huge economic consequences, a lot of anger – to the general public it seemed that a lot of players in the financial industry were fraudsters or incompetent, or both, see for example Boddy (2011) emphasising this view. That might be true in some cases, but there must be deeper reasons for the crisis – it could be that in today’s world we have much more uncertainty than we (want to) believe (Pollock, 2018). If so, the tasks in the financial market are far more difficult than expected, thus even honest and knowledgeable people cannot always avoid big errors.

However, one should not take the easy way out by claiming that the crisis was totally unforeseeable. In particular, the critical assessment of credit derivatives could rely (or could have relied) on plenty of past experience from another area – their structure is essentially non-proportional treaty reinsurance and both markets have a lot in common: the risk transfer works via pooling and layering, the data available to evaluate the risks are typically scarce, the probability distribution of losses/defaults is typically skewed, risks sometimes change rapidly over time, etc. Thus, the financial instruments aimed at transferring big insurance or credit risks are often very difficult to assess, that is their expected loss is unclear. As a result, the number of players in the market is rather limited (at least the number of institutions that “rate” the transferred risks).

This situation leads to a very special mixture of mathematical/statistical and behavioural challenges. Reinsurers have been struggling to cope with these, not always successfully, but they have learned many lessons over the course of more than one century (Section 1.1.2 of Schwepcke (2004)) in business and have above all found their way to live with a great deal of uncertainty. Some cultures established in the insurance and reinsurance world could in the future possibly inspire markets like the credit derivatives market. However, it is just as important to understand in what way the markets are fundamentally different.

1.2. Scientific Context

Many aspects of the GFC were addressed in a timely manner by experts of various areas, covering topics ranging from very sophisticated mathematics (e.g. how to deal properly with copulas – see Donnelly and Embrechts (2010) who provide besides a comprehensive list of references) to regulation issues, for example how to enhance the stability of the banking system, see Hellwig (2009) and Sinn (2010). More recent literature, such as MacKenzie and Spears (2014a), looks more closely at organisational issues, and ultimately at social interaction; interestingly, a part of this literature was initiated and/or produced by the (predominantly quantitative) actuarial profession, and in particular in the UK, see Haddrill et al. (2016), Frankland et al. (2014), Tredger et al. (2016).

1.3. Objective

The aims of this paper are twofold. We closely look at chains of consecutive risk transfer, comparing and contrasting the respective risk transfer approaches of banking and (re)insurance.

Further we draw the attention to the everyday situation of the individual practitioner in the financial industry, namely to the interaction of uncertainty and human behaviour: on one hand it is often very difficult to find an adequate mathematical language for complex economic issues and to produce reliable statistics from the limited data you have available, which makes decisions during the modelling process very delicate. Based on the (somewhat uncertain) results of these models, situated further in a very dynamic (thus somewhat uncertain) environment, the hard actuarial decisions are then followed by potentially even harder underwriting and management decisions. On the other hand, the overall uncertainty makes it very difficult to appraise, even with hindsight, the quality of the decisions made, that is whether or not they were prudent and responsible. As we will see, the resulting lack of control (and self-control) tends to push decision makers somewhat towards the risky side.

1.4. Outline

Section 2 of this paper explains what risk transfer via credit derivatives and non-proportional treaty reinsurance have in common. Section 3 describes the risk transfer chains that evolved in both markets and the troubles they caused, revealing among many shared features some fundamental differences between the two markets. Section 4 deals with skewed distributions and scarce data, a combination challenging not just quants (actuaries and the like) but also management. Section 5 narrates from practical experience how some uncertain deals come about and how quants struggle with various temptations they are exposed to at work. Section 6 concludes with lessons learned from the GFC – or partly learned some 25 years earlier. As an epilogue, Section 7 raises some questions about rules for conservative investing. For ease of reading, mathematical details are deferred to Appendix A.

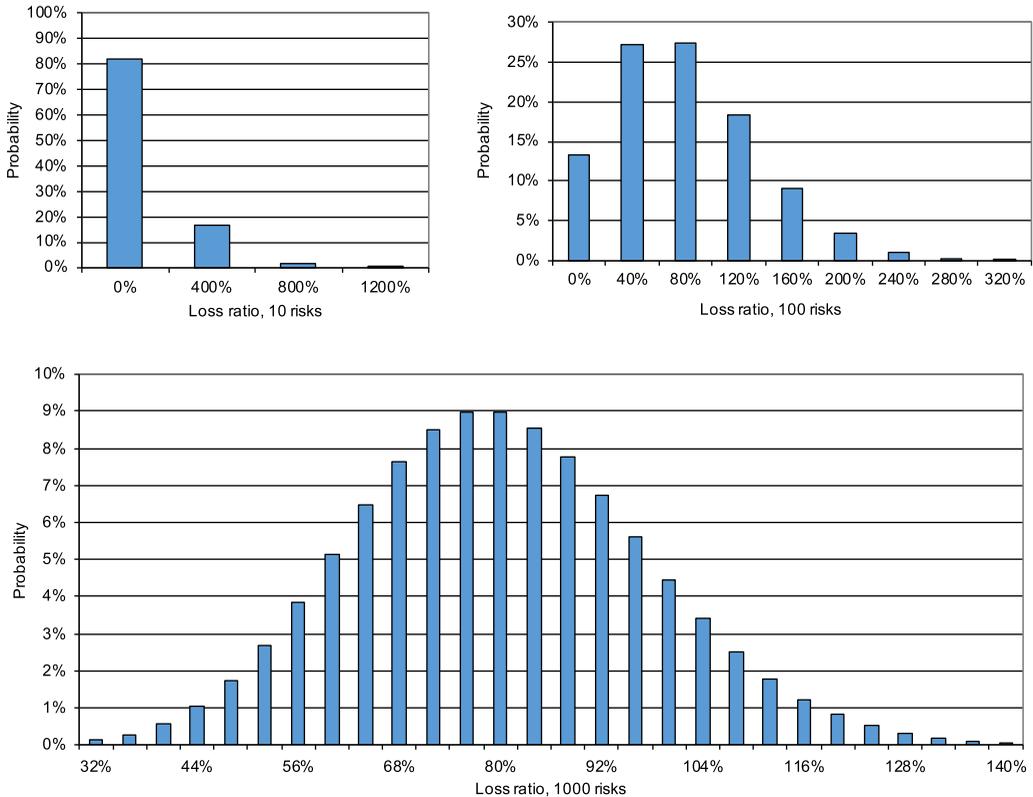


Figure 1. Loss ratio distribution for a pool of risks; average is 80%.

2. Common Features in Credit Derivatives and Reinsurance

The common essence of risk transfers via credit derivatives and treaty reinsurance is: diversify by pooling risks, reduce the pool’s probability of loss by layering. In order not to miss the subtle differences between the two markets, we describe this well-known procedure in detail.

Pooling: Risks are pooled in order to create diversification. It is well known that if the risks are not perfectly positively correlated, the outcome of pooling is less volatile than that of the individual risks. The danger of greatly overstating this diversification effect has been discussed in the aftermath of the GFC (Donnelly & Embrechts, 2010; Duffie, 2008). However, perfectly correlated risks are rare, thus pooling is in principle useful. Figure 1 shows an illustrative example (for details see Appendix A.1) for the loss ratio distribution of a pool of increasing size, from extremely skewed to bell shaped.

Layering: The aggregate risk of the pool is split by a so-called *non-proportional risk transfer*. If the loss of the pool is X , the simplest option is to transform the risk into the so-called *first risk*, $\min(X, d)$, up to a maximum d , and the *second risk* or *excess*, $(X - d)^+$. The former is “better” than X in terms of coefficient of variation (and other criteria), the latter is much more “unbalanced” than X but usually has a much lower probability of loss (Albrecher et al., 2017). The latter kind of risk may be preferred by financially strong market players as the premiums paid for such risks normally yield a higher profit margin. Furthermore, rare loss payments require far lower administration expenses. Different risk preferences have led to the so-called *layers* (in banking the French word *tranches* is more common) where the excess is cut into slices in a way that is explained by the following example from industrial insurance, being very similar to treaty reinsurance (apart from insuring large industrial risks, not whole insurance portfolios).

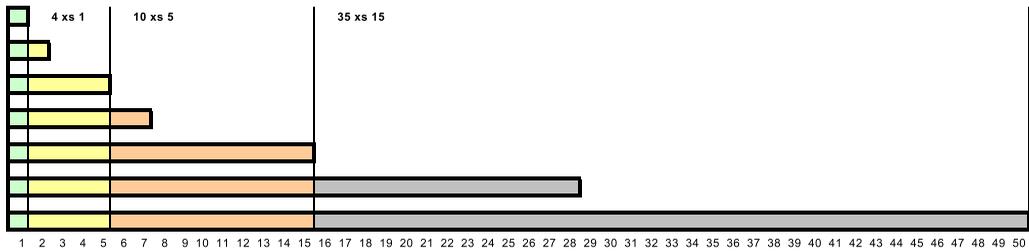


Figure 2. Allocation of losses to layers.

Suppose a large property is to be insured against fire; losses could be as high as 50 (say million British Pounds). The owners are financially strong enough to retain losses up to size 1, so they need a capacity of 49 in excess (xs) of that, shortly written: a cover 49 xs 1. They find an insurer willing to write the risk but only being able to bear a maximum amount of 4, that is they get an offer for an *XL* (= excess of loss cover) 4 xs 1, which means that they must find further insurance on top of 5. Another insurer preferring layers having a quite low probability of loss steps in, agreeing on a 10 xs 5 layer. To insure the whole risk, a further cover of 35 in excess of 15 is needed, which could be placed with a specialised insurer offering large capacities but only accepting layers having a very low probability of loss. In this way the risk is split into four parts: the first risk, usually retained (otherwise we have a four-layer program starting with a 1 xs 0 cover); a low *bottom* layer, being much more unbalanced than the risk itself but not on the extreme side; an intermediate layer; and finally a very unbalanced *top* layer. All four partners participate according to their risk preferences. See Figure 2 for how losses are allocated to the layers. In practice each layer can be proportionately split among various insurers, some of which may write shares of several layers, see Section 1.2.1.8 of Schwepcke (2004).

In treaty reinsurance (Swiss Re (2002), Section 3.3.3 of Schwepcke (2004)) this basic layer structure is applied to a variety of situations, according to how the loss is defined:

- Single loss from an individual risk in the insured portfolio: *Per Risk XL*.
- Accumulation of all losses caused by a catastrophic event (e.g. hurricane, earthquake): *Cat XL* (Accumulation XL).
- Aggregate of all losses occurring in a year (in a line of business): *Stop Loss*.

In order to make layered (re)insurance easily comparable to credit derivatives, let us interpret these layers as simply financial instruments creating certain cash flows. Note that in this paper we only treat traditional reinsurance – Financial reinsurance can have quite different cash flows.

- The insurers of the top layer receive a premium every year and in most of these years they will not have to pay anything. Due to the very low probability of loss, the premium is only a small percentage of the cover. Thus, they get a return that is steady but small compared to the cover (the capital at risk). In the very rare (say once in 100 years) case of a loss, in particular if it is a total loss, they will have to pay a high multiple of the annual yield. This cash flow has much in common with investment grade bonds, which have a low annual return (being usually defined as spread over the risk-free interest rate) compared to the invested capital, and in the very rare case of default the loss is as large as the return accumulated over decades.
- The bottom layer receives a premium every year, being a higher percentage of the respective cover due to the higher probability of loss. There might still be loss-free years, but losses are not exceptional events. If we compare this to bonds, the most similar type would be junk bonds, having very high returns compared to the invested capital, while on the other hand bearing a considerable default risk.

- Intermediate layers are somewhere in between, in the same way as bonds having an intermediate security.

Now we explain credit derivatives (see e.g. McNeil et al. (2005), Duffie (2008), Gorton (2009), Hellwig (2009)), namely the simplest case of an ABS (Asset Backed Security). This is created out of a pool of mortgages (or loans, etc.). What corresponds to the insured loss is the difference between the money (principal plus interest) that contractually has to be paid back by the borrowers and what is actually paid back, being less if some mortgages default. The cash flow is different from insurance in that here the whole money at risk is handed over in advance and then gradually paid back until redemption. We will see soon how important this difference is. The layering works as follows: the aggregate amount expected to be received from the borrowers is sliced into tranches, which are given different priorities for the incoming cash flow. Three layers are a minimum but there can be more. The bottom layer is called the *Equity* tranche, it has the highest (nominal) return and the highest probability of default. The intermediate layer is called the *Mezzanine* tranche (adopting a term for an intermediate floor). The top layer is called the *Senior* tranche; it has the lowest return but the (estimated) probability of loss is so low that it gets an investment grade rating. The incoming cash flow from the borrowers at first goes wholly to the Senior tranche until its amount due is settled, then the cash flow is redirected to the Mezzanine tranche. Once the latter has received what it deserves, the remaining payments (if any) go to the Equity tranche. This order of priorities ensures the desired order of probabilities of default. Formally, each tranche is a separate bond, assigning to the respective investors the right to get the payments from the borrowers according to the defined rules.

When this kind of derivative evolved it was largely expected that, as in reinsurance, the originators would retain the Equity tranche – which would have been an effective incentive for them to carefully select the borrowers (Hellwig, 2009). In the end, it turned out that there were investors such as hedge funds willing to buy such tranches (Akseli, 2013), just as there are investors buying non-rated junk bonds.

If we compare ABS tranches to insurance, they are closest to Stop Loss reinsurance treaties. The latter insure the aggregate loss of a certain portfolio of insurance policies in a certain period, while the former “insure” the aggregate default resulting from a portfolio of mortgages.

For completeness it shall be mentioned that in both areas, insurance and credit derivatives, there are various refinements of the explained layer structure. For example, reinsurance layers have a variety of features working essentially as loss participations (Albrecher et al., 2017; Section 6.3.3 of Schwepcke (2004)), credit derivatives may treat interest in a somewhat different manner than principal (Gorton, 2009). However, these features do not change the basic characteristics of tranches of different priority and probability of loss, so we leave them aside here.

From an actuarial point of view, it is clear that while it is usually not too difficult to estimate the risk premium for a portfolio of insured risks or the adequate coupon for a pool of mortgages, respectively, by assessing the average frequency and severity of losses/defaults, far more is needed to associate the correct portion of this risk premium to each layer/tranche. This requires the calculation of the distribution function of the losses/defaults, that is far more complex models taking into account in particular the dependency between the pooled risks, which in turn might require far more data. We shall revisit this point the following three sections.

3. Risk Transfer Chains

In order to optimise its risk profile, it is straightforward for an insurer, or an investor in credit risks, to think about transferring a part of the assumed risks to other parties. In fact, this happens in both markets.

3.1. Structure

For the (somewhat diversified) portfolio of an insurer there is non-proportional reinsurance available, protecting against fluctuations due to exceptionally high losses (Swiss Re, 2002). A reinsurer writes hundreds or thousands of such layers, creating a (somewhat diversified) pool. To protect this pool, the reinsurer can in turn buy a program of layers. Reinsurance for reinsurers is called *retrocession*. There are specialised companies doing this kind of business; however, some reinsurers act as retrocessionaires, too. A retrocessionaire can in turn look for XL protection of its (somewhat diversified) book of business, and so on. This results in a risk transfer chain where each step consists of pooling and layering:

Risk → Insurer → Reinsurer → Retrocessionaire → 2nd Retrocessionaire ...

The same occurs in the credit derivatives market (Gorton, 2009; Sinn, 2010), in particular with the Mezzanine tranches. While it is not a problem at all to place top-rated tranches among the increasingly many investors preferring (or being restricted to) investment grade bonds (Caballero et al., 2017), and on the other hand one often finds risk-loving investors buying Equity tranches, there is apparently no substantial demand for intermediate tranches. However, by pooling, say, one hundred non-investment-grade tranches – perhaps covering different countries and different kinds of underlying risk: mortgages, loans, credit card, leasing, etc. – and slicing this (somewhat diversified) portfolio into three layers, one can create new bonds called *CDOs* (Collateralised Debt Obligations): the new top tranche should have a much lower probability of default than the average Mezzanine tranche, thus has a chance to get a top rating; the new bottom tranche is a further product with high risk and high nominal return; and in between there is a new but now much smaller Mezzanine tranche. If no one wants to buy the latter, it could again be pooled with others, and so on. This results in a risk transfer chain where likewise each step consists of pooling and layering:

Mortgage → ABS → CDO → CDO-2 → CDO-3 ...

In principle this is a good idea. Why not create tailor-made products having exactly the probability of loss that investors like most? However, one should bear in mind that each step requires the modelling of the aggregate loss/default distribution of a portfolio of risks that are arguably not completely independent of each other. Even without considering the details of such modelling (treated in McNeil et al. (2005), which was notably published before the GFC) it is clear that any model error (due to scarce data or some other reason) will propagate from one chain link to the next and further along the chain. And beyond quantification further issues may arise.

3.2. Wandering Defaults

Look at the following risk transfer chain:

A → B → C → D → E → F ...

What happens if D fails while the risk transfer is in force?

Here we have to distinguish the seemingly similar cases of (re)insurance and credit derivatives. It turns out that the different order of payments creates a fundamental asymmetry between the two markets.

In the case of reinsurance, if a loss hits various chain links and D is unable to pay its part, this firstly matters to C: C reinsures B and has to indemnify B no matter whether C has reinsurance protection from D, and whether D pays or not. In other words, C bears the risk of default of its reinsurer D – this credit risk is an incentive for selecting financially strong reinsurers. Only in case the failure of D causes the insolvency of C, the parties further to the left (B and possibly A) are affected by D's bankruptcy. E and the further parties to the right are not affected at all; they will

have to indemnify their part of the loss independently of the insolvency of D. In short, if insolvencies spread along the chain, they spread to the left.

Consequently, if you are a reinsurer thinking about joining a chain by taking and ceding risks in the described manner, you have good reasons to look both to the left and to the right, that is to carefully select the parties to take risk from and to cede risk to, respectively, as bad events can come from either direction: from the left you assume underwriting risk, from the right you assume credit risk. The resulting both-way cautiousness should generally keep weak players out of retrocession chains, which ensures the stability of such chains and keeps them generally rather short because the number of available market players being both professional and financially strong is limited.

Let us now look at credit derivatives chains. Recall that in an ABS/CDO deal the whole money at risk is handed over at first, to be (most probably) returned later. Thus, during the creation of the chain every participant pays a large amount of money to his predecessor (acquisition of the bond), then over the years this money (plus interest) gradually flows back from the left to the right starting from the initial borrowers through the whole chain. If now D defaults (partially or fully), the payments to E are interrupted or reduced, which can create a domino effect to F and to the further parties to the right. In short, if defaults spread along the chain, they spread to the right.

This completely changes the characteristics of the chain. As in the reinsurance case, each player entering a chain should look to the left where the risk is taken from. But here you have no incentive at all to look to the right, that is to carefully select the people you are selling CDOs to: as soon as they have acquired the bonds, the only link to them is that they are entitled to receive payments from you, thus their financial situation does not affect you any longer. Summing up, in CDO chains bad events can only come from the left. There is no risk in the cession (except for maybe a reputational one); you can cede to any party wanting to take the risk, be they financially strong, professional, or neither.

What makes this situation even worse is the availability of potential risk takers: in hot market phases, when the demand for investments is very high, the originators of credit derivatives are able to transfer 100% of the risk they have assumed (*originate to distribute*), by selling all tranches completely and at favourable prices to investors, be they credit market experts or not. In this situation a market player is not even incentivised to look to the left, that is to carefully evaluate the assumed risks – these will anyway be repackaged and sold within a few days. Indeed, this largely occurred in the years prior to the crisis (Duffie, 2008; Hellwig, 2009).

The fact that most investors need to base their investment decisions on ratings from recognised rating agencies, together with the time pressure of the booming credit derivatives market, led to situations like the following, being reported by practitioners: while the originators were able to dedicate considerable time to the pooling and layering of credit risks, employing complex modelling tools, which was followed by the assessment on behalf of more than one rating agency in order to confirm the aspired rating, afterwards the newly designed bonds were marketed by brokers who gave potential investors not more than ten minutes (!) to decide whether to take or to leave the deal. In the end investors willing to buy this kind of product had no chance but to rely completely on the assigned ratings, that is the assessment of the investment was outsourced to the rating agencies without double-checking it by thorough in-house analysis. It goes without saying that ten minutes is not even sufficient time to do the non-mathematical part, that is to read the documentation about structure and content of the preceding chain links.

3.3. Cycles

As we have seen, risk transfer chains can be problematic in their own right – but the really bad variant is as follows. The parties C and F could be the same. This can happen to all players ceding and writing risks of the same kind. If they always read all the documentation about the

risks to be assumed, such cycles should be avoided, but it is clear that when having to check a long series of pools and layers some details can be missed and so inadvertently you become your own reinsurer.

Old reinsurance practitioners occasionally tell stories like this:

- You have to pay a large loss – you just got a 50 million US Dollar cash call from one of your clients. Now you check your retrocession treaties. To your great relief, a “retro” layer covers the loss after a retention of 5 million. Happily you send a \$45 million cash call to your retrocessionaire.
- However, to diversify your portfolio, you have written a bit of retrocession business. A few weeks later, your retrocedent (the company you protect) sends you a \$38 million cash call, referring to the above loss event.
- If the retrocession business you have written is protected by your retrocession treaty, you can prepare your next cash call and then wait: sooner or later you will receive a further cash call.

It is truly a challenge to calculate the final share of this loss event that you, and all the others involved, will have to pay.

What makes things even worse is the fact that chains are not just chains – every risk transfer step means pooling (and connecting) a hundred or more market players. The result is a kind of spider’s web linking everyone to everyone through chains and cycles. If, further, some of the people in this web go bankrupt, things start getting really complex . . .

This is not just a funny story about some minor market players. As was reported even in the mass media, a considerable part of the GFC was exactly this: an all-embracing web of CDOs and other credit derivatives trapping investors worldwide, with defaults, downgrades, and liquidity problems infecting other players, and so on. It must be noted that beyond layered risk transfer there were further critical links between players, such as commitments of liquidity assistance (Hellwig, 2009), plus the overall correlation of the market prices of credit derivatives. To the general public this phenomenon may have seemed unheard of, but it was not the first big event of its kind: reinsurers still remember well what happened to the Lloyds of London market in the 1980s – it is called the *LMX (London Market Excess) Spiral* (O’Neill et al., 2009). Focusing on what matters for our topic, in that period Lloyds, in order to enlarge the number of players in their market (“Names”), greatly eased access conditions and acquired a lot of new members. At a certain stage, there was a surplus of insurers. To supply the increased demand for business and to take advantage of the resulting soft-market conditions, a lot of players began to retrocede a lot of their written business via XL retrocession. This process was strongly pushed by brokers – they got a considerable commission for every placed deal, no matter whether or not it was a useful risk transfer. As long as no large losses occurred, everyone seemed to be making money. Then the Piper Alpha oil rig exploded, causing a huge market loss involving many parties. It was followed by further large claims, which drove several Names into insolvency and made clear that the only ones having made profit from this business model were the brokers.

Lloyds survived their crisis, but it was a huge task. They created an appropriate entity, *Equitas* (Section 1.6.2.1 of Schwepcke (2004)), to manage run-off and correct allocation of the losses of that period.

3.4. Tail Geometry

Let us draw attention to a big problem arising when layers protect (pools of) layers, due to the typical geometry of loss size distributions. We explain this “XL on XL” situation with an example from insurance. Say a portfolio produces 300 losses per year. How many of these exceed 1 million Euro? In practice, it is often only a tiny part, for example on average 0.3 per year, which is

1,000 times less than the “ground-up” number of losses. What is the frequency of losses exceeding 2 million Euro? It could be 1,000 times less. The frequency at 3 million could again be 1,000 times lower than this, and so on – in the case of an exponential severity distribution we would have exactly this situation. However, in practice, in the million Dollar range one often observes much heavier distribution tails than exponential, in particular Pareto-like ones (Schmutz & Doerr, 1998), which can lead to surprisingly low decreases of the loss frequency. For example, if the severity has a Pareto distributed tail with parameter $\alpha = 1.5$, more than half of all losses exceeding 2 million also exceed 3 million, and most of these exceed 4 million, etc. For exact figures and technical details see Appendix A.2.

Suppose there is a 4 xs 2 (million Euro) reinsurance layer protecting this portfolio and the reinsurer writes 99 alike 4 xs 2 layers covering other businesses. The pool of these 100 layers shall be protected by a retro cover 200 xs 200. Suppose that each layer limit 4 holds per loss and per year, such that the retro cover fits the maximum possible pool loss 400. How much of the aggregate premium of the 100 layers has to be paid for this XL protection of the pool? We focus on pure premiums here, that is expected loss. The two extreme cases are the following:

1. If the outcomes of the 4 xs 2 layers are stochastically independent, an aggregated loss of more than 200 million Euros is extremely unlikely – to exceed this amount more than 50 independent layers need to suffer a loss in the same insurance year. See the detailed calculation in Appendix A.3. Thus, the risk premium of the retro layer would only be a negligible part of the pool’s aggregate premium.
2. If the 4 xs 2 layers have identical severity distributions, namely according to a Pareto curve with $\alpha = 1.5$, and if, further, the respective losses are comonotonic, the retro layer suffers a loss if each 4 xs 2 layer is affected by a loss being so large that the layer has to pay more than 2. Situations where each layer suffers two or more smaller losses may also trigger the retro layer, but for our purpose it is sufficient to focus on the large-loss case. If we artificially split each 4 xs 2 cover into two layers 2 xs 2 and 2 xs 4, the retro layer covers all losses of the 100 2 xs 4 layers, thus its risk premium equals or exceeds the sum of the premiums of these artificial layers. As the layer severities are heavy tailed, the 2 xs 4 layers are not much cheaper than their 2 xs 2 counterparts, thus require not much less than half of the premium of the original 4 xs 2 layers. Indeed, for the given Pareto tail the premium share of the 2 xs 4 layers is 31%, see Appendix A.2. Overall the adequate premium for the 200 xs 200 retrocession is about a third of the total pool premium – an astonishingly high figure, which is due to the missing diversification effect in this XL on XL case.

These two dependencies are extremes – real-world cases are mostly somewhere in between. However, it must be noted that although comonotonicity is very unlikely, cases of strong dependency can produce similar results. Many people, when seeing a layer starting at 200 million Euros, think it must be very unlikely to suffer a loss. But, the above example illustrates that if heavy tails and strong dependency come together, the probability of such events may be hundreds of times higher than expected.

The extremely wide range of possible outcomes of XL on XL deals makes clear that there can be an enormous uncertainty in model selection and parameter estimation, which furthermore increases with the length of the chain due to the “accumulation” of XL on XL situations (Donnelly & Embrechts, 2010).

Although model risk is being increasingly discussed, it seems that it is still popular to interpret bad surprises as random events (“very bad luck”): during the GFC some bankers claimed they would have experienced “several 25-standard deviation moves in a row” – an event as rare as winning the lottery in more than 10 countries on the same day (Dowd et al., 2008). However, what they experienced was certainly a consequence of model uncertainty. In other words, the applied models were fatally wrong.

3.5. Workarounds

Reinsurers have been facing XL on XL situations and risk transfer chains for some time and have accordingly established some rules and traditions to reduce the likelihood of very bad situations:

- To begin with, 100% risk transfer is close to impossible in reinsurance (apart from special cases where the reinsurer is in control of the underwriting process). The dominant philosophy is *originate to hold*, or as practitioners say: “If a company does not want to retain even a tiny part of a certain business, they must be very afraid of something – whatever it might be, don’t cover it.” A retention of a part of the risk aligns the interests of ceding and risk-taking party in a natural way.
- Reinsurers traditionally calculate the adequate premium for the risks they assume themselves, that is they barely rely on premium ratings from external parties (apart from adopting some external expertise about issues like earthquake probabilities). They would typically be involved in the structuring of the reinsurance (insurer’s retention, etc.) or check at least whether the proposed structure makes sense.
- They routinely conduct accumulation control in order to limit their exposure to certain named large individual risks and accumulation events, which traditionally includes natural disasters but nowadays also terrorism scenarios, for example. That means that if a company notes they are about to exceed the given limit for say hurricanes in Florida, they would stop writing Property reinsurance in the area – or look for an adequate transfer of a part of the risk.
- Reinsurance treaties traditionally exclude excess business, that is the reinsured policies must not be insurance layers. Only policies with rather small deductibles are admitted. If excess business is not excluded, total transparency is required: the insurer has to provide a bordereau, that is a list of the individual layer policies with all information needed to do a sound XL on XL rating.
- To avoid cycles, reinsurance treaties usually don’t protect reinsurance business; retrocession treaties may exclude retro business.

Of course, there are situations where full transparency by the ceding party would be undue. In particular, retrocession tends to be more “anonymous” than reinsurance. However, in the important field of Accumulation XLs for natural catastrophes there is a fair workaround: one can refer to figures from geophysical simulation models being known and accepted throughout the industry. Such model output yields an assessment of the exposure per geographical area without having to disclose the individual risks.

And then there are situations where lack of transparency is knowingly accepted, maybe between parties having a long-term relationship including the commitment to let the other party recoup somehow in the years following an unexpected large loss. Alternatively, one can find retrocessionaires of the more risk-loving variety, asking very few questions, being aware that what they do is essentially gambling, and playing the game. It is possible to do business which such people, however, one should know the rules. As practitioners put it: “You can be certain that this market will triple premiums as soon as it suffers a loss.”

Apart from the exaggerations of the LMX spiral, the typical risk transfer chain in reinsurance has not more than four to five chain links. People in the industry seem to be well aware that longer chains have such a high model risk that their added value is questionable. In the credit derivatives market, though, there are rumours about some far longer chains (Sinn, 2010), which were created to cater for the huge demand for investment grade bonds in the years preceding the crisis.

Some of the measures developed by the reinsurance industry to deal with risk transfer chains might be not too difficult to adapt to markets like the credit derivatives market, in particular those preventing cycles and transparency standards. A mandatory retention of part of the assumed risks

was recommended early by many experts, for example Hellwig (2009), and has to some extent been implemented after the GFC. Very simply, big markets like Europe and the USA have introduced a minimum 5% retention, which can be complied with in a number of ways – and which will likely have to evolve further to be as effective as desired (Akseli, 2013; EBA, 2014; Krahnen & Wilde, 2017). Arguably much harder to address is the lack of incentive in credit derivatives chains to carefully select whom to cede risk to. Maybe it is worth thinking about whether it would be possible to implement the certainly very useful pooling/tranching system in a more insurance-like fashion.

4. Skewed Distributions and Scarce Data

Even when not very heavy-tailed, distributions in insurance are typically skewed towards the “bad” end, and the same is true for mortgages/loans – gains are rather limited and losses are potentially huge. To assess such distributions, one typically needs more data than for a risk whose outcomes follow a nice bell curve, and often there are less data available than one would feel comfortable with. The decisions to be made in such situations are not only a statistical problem but also a behavioural one, as can be seen from the following (fictitious but not unrealistic) examples.

4.1. The Interpretation Problem

Imagine you are a quant having to look at a business having a skewed loss distribution. Very simply, skewed means that if you observe such a risk for, say, fifty years, you typically see one very bad year, some years may record slight losses, but most of the years yield very nice results. However, in practice, you rarely have more than ten years of loss history. If older data exist, you can often not use them due to changes in database structures or because it is felt that the world has changed too much since then to rely on that data. So, you are in one of the following two situations:

Case A: There is *no very bad year* in your data.

This loss history looks stable (maybe nearly normally distributed). Modelling should not be too difficult. You wonder whether the data are representative. *Yes, of course!* Everyone agrees about that: the originators of the business, the broker trying to place the risk transfer with your company, and your colleagues being keen on creating business volume.

Case B: There is *a very bad year* in your data.

This loss history is very unfavourable for modelling purposes. Can this data be representative? *Of course not!* That bad year was without doubt a very exceptional event, having a return period beyond 200 years. Everyone agrees about that: the originators of the business, the broker trying to place the risk transfer with your company, and your colleagues being keen on creating business volume.

It is clear that doing a good job in such situations challenges not only the technical skills of a quant. More mathematically, if the loss distribution for a single year is skewed, this persists for periods of about 10 years. Of course, the latter distribution is less skewed, but usually still far from a symmetric bell curve. See Figure 1 for how slowly a growing pool approaches the normal distribution: the same holds if we observe a “skewed” risk for an increasing number of years. That means that most of the typical statistics you come across when assessing such risks show a better result than the long-term average. If you ignore the few really bad years you happen to observe in your ratings, or assign excessive return periods to them, your ratings will on average be too cheap.

This problem is in principle well known and as a consequence people work with parametric loss distribution models able to account for issues like skewness. However, in the practice of non-proportional risk transfer things are not that easy. Most statistical methods have been developed for large amounts of data. They can be applied to small data sets, too, but with skewed underlying distributions strange things may happen, as Section 4.2 illustrates.

4.2. The Calibration Problem

Imagine again you only have data from the past ten years available but need to calculate the 100-year event, that is the 1% quantile (maybe in order to rate a high layer or for a VaR calculation). You start your powerful statistics software, which provides a bunch of loss severity models (see Appendix A of Klugman et al. (2008) for an overview) and is able to select the best fit according to well-established statistical methods.

Tool output: best fit is *Weibull*, $VaR = 100$ (say million British Pounds).

The day after you get updated data (slight changes due to run-off). You rerun the calculation.

Tool output: best fit is *Lognormal*, $VaR = 50$.

Later that day you get a new data update (someone found a minor typing error). You rerun the calculation.

Tool output: best fit is *Lomax* (a Pareto variant), $VaR = 150$.

What do you do now? If this result is not just the premium rated for a minor deal, but an important figure for, for example, solvency purposes, what do you tell management? They probably don't want to hear stories about best fit procedures yielding surprisingly unstable results (Jakhria, 2014).

This example is a bit stylised, but it describes the uncertainty inherent in the modelling of skewed distributions with limited data in a realistic way. See Appendix A.4 for the detailed calculation of a similar case. Indeed, high uncertainty can affect quite simple settings: here we just have a one-dimensional fit and few parameters to estimate, which is a far simpler situation than the complex copula models (and the like) needed to rate risk transfer via pooling and layering in the credit derivatives and retrocession markets. It is clear that in such complex situations even 30 or 50 years of representative data cannot remove such uncertainty – model selection and parameter estimation will be highly sensitive anyway.

What makes things still worse is the dynamic environment both in the reinsurance and credit markets. Earthquake activity may be stable over long time periods, but other natural hazards may change more quickly, as do social and economic conditions. A paramount example is US house prices (Sinn, 2010), which for decades altered between upwards and sideways movement – the first significant setback in 2006 triggered the credit derivatives crisis. The dilemma is as follows: either you use only the data of the very recent past (then you have few observations and are likely to observe a period that is better or worse than the long-term average) or you include older observations (then you have more data and likely embrace different kinds of periods, but have to incorporate the shifting environment somehow into your models, which makes them more complex, requiring in turn further and more detailed data).

In short, whatever methodology quants adopt, when the available data are not abundant, they will have to make some hard decisions. And some colleagues have a similar problem.

4.3. The Management Problem

Companies need to be protected against very bad years. That is the classic situation where management come across skewed distributions – without precautions the economic results of the

company will have a distribution being skewed towards the bad end. Essentially, management have the same problem as a quant having to do a rating based on 10 years of data. The recent past in most cases has not experienced a bad year, the remote past doesn't seem to tell anything as "those were totally different times". Of course, some plausible bad scenarios have been developed, but no one can tell for sure whether these are 50-year events or far less likely.

One thing is clear: if management, while the company faces cost pressure, has to choose between two strategies to reduce the risk of the enterprise, the less effective strategy is appealing – it will be cheaper (at least in 49 out of 50 years). In such a situation, it is very hard to stay risk-averse and to enforce the safe option.

There is no point in playing management and quants off against each other. The uncertainties they have to deal with make it very difficult for both professions to do a good job. However, this is no excuse for negligence.

4.4. Workarounds

Reinsurers have developed culture to address the problems arising from skewed distributions and scarce data. Their advantage over capital markets is that a lot of their business relationships last for decades and involve several reinsurance treaties, which are revised and renewed every year. This firstly gives them the chance to negotiate a premium level that both parties regard as a fair long-term average, even when they don't agree about every single deal. In a way they are able to smooth several uncertain premium calculations across time and business lines. Such long-lasting ties secondly lead to data series far beyond 10 years, which enable reinsurers to learn – slowly, but steadily – over time which of their assumptions came true and which did not. Apart from being a base for continuous improvement of rating methods, this might also have provided reinsurers with some intuition about which stories to believe and which not to believe. So, if a huge loss has occurred in the past 10 years and the ceding company or the broker assures: "That will never happen again in this portfolio", they will remember cases when they had heard exactly that and a similar loss did occur just a few years later. Such experience possibly creates the right portion of scepticism needed to avoid some of the very bad deals.

Generally, long business relationships help reinsurers to keep in mind the remote past. This is not only a matter of data availability – tradition helps, too. Reinsurance became a big market due to its decisive role in various rare events, starting with the 1906 San Francisco Earthquake, see for example Swiss Re (2002). It is natural for the industry to look back very far and this attitude helps both actuaries and risk management to be listened to when they warn against bad events.

The calibration problem is arguably most difficult to address. The word *calibration* is often used to describe a rather automatic and standardised statistical inference process. Such procedures certainly make sense when a lot of data has to be processed, say for a weather forecast. However, the above calibration example makes clear that the results of statistical procedures are basically random outcomes, and in cases of scarce data very volatile ones. Powerful statistical tools cannot solve this problem. Reinsurers have adopted what is proposed in various publications about Extreme Value Theory (Embrechts et al., 2013; McNeil et al., 2005): inference of heavy tails is basically a step-by-step procedure requiring many intermediate decisions, successively employing various methods from deep exploratory analysis to parametric models (though using possibly non-standard estimators having good small-sample properties, see Brazauskas and Kleefeld (2009)), and one should always test the sensitivity of the model assumptions and the plausibility of the final results. It might be more difficult to stick to such time-consuming procedures in the hectic placement of capital market deals; however, in the long run, slow and careful calculations are the cheaper option.

In the following, we will see some consequences of uncertain calculations and their very random results, being in part far too expensive and in part far too cheap. Note that this problem is not restricted to fancy businesses and overly complex risk transfer deals. Even in very traditional

and professional markets, it occurs that the data are far from sufficient to fairly rate a deal, for example, when new insurance lines come into existence, or new insurers, or sometimes even new economies or states. They need (re)insurance, loans, and other financial services, just as business areas providing abundant and reliable data.

5. Decisions in the Face of Uncertainty and Incentives

It is easy to make decisions, and to stand up for them, if you have plenty of information. Under a lot of uncertainty this is much harder. Say you rate a deal and the data are so scarce that you need to make an assumption (to use judgment to choose a model or a parameter). Say you have two scenarios, both looking plausible, being not much different from each other – and the more conservative one leads to a three times higher premium. Data cannot tell you which assumption is more realistic (and colleagues cannot either). In this situation, there are pressures of various kinds pushing you towards the cheaper option. The importance of behavioural issues, in the face of uncertainty, to organisations in general has been recognised for a long time – see, for example, Kahneman and Lovallo (1993) – and lately also by the banking (Pollock, 2018) and the insurance industry (Frankland et al., 2014; Haddrill et al., 2016; Jakhria, 2014; Tredger et al., 2016; Tsanakas et al., 2016; Weick et al., 2012). Many of the issues described in the following appear indeed in more than one of these references.

By way of illustration, Section 5.1 starts with a story, as practitioners in reinsurance tell it (but it might also occur in other areas).

5.1. The Deal-Placing Game

When a “bad-data” deal is offered, some players in the market give it a chance and have a closer look at it. And then the same series of events tends to unfold:

- With so many hard decisions to make, time is never sufficient.
- People get hectic, with lots of discussions and meetings.
- The frequency of the broker’s telephone calls increases steadily.
- At a certain point, a market player offers a share in the deal, (unknowingly and inadvertently) at an arguably too cheap price.
- The broker is happy, gives that company the lead in the deal (say they take a 30% share) and attempts to place the remaining 70% at the indicated price.
- At this point other players, who were thinking about charging much more, hope that the leader knows what they are doing, and take a minor share in the deal.

This is basically herding behaviour – someone being supposedly knowledgeable leads the way, and others follow. At the same time, it is a variant of the well-known winner’s curse.

The hazard is obvious: if many deals are placed this way, the written business will on average have a too low price, so the market as a whole will lose money. But it is so hard to say no – after all the offered price is within the range of plausible assumptions. Furthermore, if you go for the cheaper option, no one will be able to prove it was a wrong calculation. And finally, the fact that others in the market accept the deal is a kind of rating classifying the deal as acceptable. Even if you are generally sceptical about external ratings, in such a situation they are difficult to argue against.

What you experience here is a combination of double peer pressure (by the people around you being keen on business and by the players having already accepted the deal) and lack of control of your work (due to the high uncertainty). This is a threefold incentive to stay a bit on the cheap side.

The appraisals of other market players have a particularly strong weight in markets like reinsurance having a quite limited number of active market players, who essentially “know” each

other. Of course, there are a large number of less known minor players, but these typically write small shares of deals after all terms and conditions are agreed on, thus do not influence the structuring/pricing process, see Section 1.5.2.3 of Schwepcke (2004). So, if the proposed premium comes from a renowned company, how dare you say it is inadequate? “They must know that kind of business in that country!” is the thought.

While the credit derivatives market has many more players than reinsurance, at least during the credit derivatives boom nearly all of them adopted a passive role (Duffie, 2008; Hellwig, 2009): the assessment of the probability of default of the structured bonds was done by the few big rating agencies (being supported on the way by the originators); most risk takers would never question their ratings. How dare you challenge the assessment of a big company being officially in charge of the rating of these bonds? Recall that scarce data are neither sufficient to clearly confirm their judgement nor to contest it. The rating agencies had in any case the most powerful role, being essentially referees for credit derivatives. As an aside, they had vested interests (Sinn, 2010). As referees they were usually paid by the issuers of the rated bonds, thus economically depended on the issuers. Worse yet, they could play a second role, acting as consultants for the same issuers helping to structure the derivatives in order to obtain the aspired ratings. A substantial redesign of this business model has (by the early 2020s) not yet taken place, which means that the inherent moral hazard is still there.

Several variants of herding behaviour could be observed in the credit derivatives market. Investors followed rating agencies, which in turn followed mainly a certain way to apply the Gaussian copula model, the “Li model” (Donnelly & Embrechts, 2010; Duffie, 2008). For being easy and efficient to use, it became the most widespread method to assess default correlations, despite some inherent flaws, which were notably recognised (McNeil et al., 2005; Whitehouse, 2005) years before the GFC. See MacKenzie and Spears (2014b), MacKenzie and Spears (2014a), and Puccetti and Scherer (2018) illustrating why and how the Li model became so popular among practitioners. Its flaws led, in particular, to frequent underestimation of upper tail probabilities. One can suspect that the very low probabilities the model assigned to high tranches helped boost its popularity.

Sticking to what appears to be a “market standard” is tempting, particularly in uncertain situations. How many debates can you spare if everyone agrees about the methodology – you can justify what you are doing by simply saying “this is common in the industry” (Haddrill et al., 2016). From a risk management perspective, however, this behaviour constitutes a *concentration risk* for the market as a whole.

A further kind of herding behaviour was investors’ overall view of senior CDO tranches. At a certain point it had become doctrine that these were the best business opportunity of the decade, such that even sceptical investors could barely resist them (Hellwig, 2009).

Reverting for a moment to the above deal-placing game, occasionally there seem to occur really bad variants, as practitioners report: for example, someone claims (wrongly) that the proposed premium was quoted by a certain renowned market player. In this or similar ways various reinsurers (or maybe rating agencies?) can possibly be played off against each other – or, as an extreme case, various colleagues of the same department. People simply shop around until they get all the signatures they need.

All in all, there are various incentives to avoid very expensive calculation results:

- Human beings tend to go with the crowd. If you accept the deal, everyone around you is happy (at least in the near term).
- An assessment without calculating many variants and asking many questions saves a lot of time. The more time you dedicate to a single deal, the more other work will be waiting for you afterwards.
- Admitting uncertainty is precarious. If you try to discuss the case with colleagues, this might harm your reputation – furthermore, they are too busy to spend much time on your problem.

- It is even hard to admit to *yourself* that you are very uncertain, that is to acknowledge that all the skills you have acquired and all the hours of hard work are insufficient to carry out a certain task with the desired degree of certainty. Humans feel much more comfortable when they believe they are in control of the situation, so in difficult moments they are somewhat vulnerable to deluding themselves. This is called the *illusion of control* (Kahneman & Lovallo, 1993).
- If you instead take the uncertainty into account and choose a more conservative rating variant, this might affect the business result assigned to your department for variable salary purposes. In other words, a rather expensive calculation could affect your bonus and/or, worse, the bonus of your boss/colleagues.
- Due to the high uncertainty, it will be almost impossible to classify your calculation as wrong. Deals of this kind generally have such a low probability of loss/default that the outcomes in the near future will most probably be positive or at most slightly negative. In the rare case that the deal produces a catastrophic result, you will get away with just claiming that this was an extreme random result (“bad luck”) and moreover totally unforeseeable.

5.2. Workarounds

It is apparently very difficult to handle the problem that calculations that are on the cheap side tend to create (in the near term) less trouble and less workload in the company. To assess the quality of the work of quants, one has to look at a lot of calculations together in order to smooth the random effects inherent in individual deals. In insurance this is called *re-pricing* – written deals are analysed later to see whether the rated premiums were (on average) too cheap or too expensive compared to the actual loss experience. This is a complicated and tedious exercise, which might be easier to implement in the underwriting cycle of a business like reinsurance, which has essentially one pricing season per year, than in capital markets placing deals throughout the year under a lot of time pressure. However, if the rating of difficult business is not given sufficient time, the variety of pressures illustrated above will push the rating results towards the cheap side.

Finally, a word about a special case of gambling comprising a special temptation: it was reported that during the GFC some players made profit from the failure of others, in particular through speculation with *Credit Default Swaps* (CDSs). These derivatives were originally designed to protect lenders and other stakeholders against the default of a certain institution, yet they can be traded on the capital market. Thus, someone having no stake in the so-called *reference entity* of the CDS could buy the derivative – and later on be induced to accelerate the default of that entity, say by spreading certain rumours in the financial market, etc. (O’Neill et al., 2009; Sinn, 2010). This problem has been known to the insurance world for a very long time. For example, someone could take out life insurance referring to the life of his neighbour. If this healthy neighbour then dies surprisingly early, was it a coincidence – or was it arsenic? To prevent the moral hazard inherent in such situations, a principle was introduced centuries ago: to be indemnified by an insurance policy, one has to prove an *insurable interest* – one must actually have suffered a loss. One could think about whether some areas of the capital market, for example CDS arrangements, could adopt this principle (Posner & Weyl, 2013). Apart from wrong incentives, CDSs and products like synthetic CDOs being based on them, which boomed in the years before the crisis, amplified the exposure to credit risk beyond the real risk of the US real estate market (Gorton, 2009). The leverage on the real mortgages was a key element of the resulting systemic disruption.

6. Conclusion

The comparison of the GFC with challenges met earlier by the (re)insurance industry identifies ways in which the credit derivatives market could possibly be enhanced. On the other hand, the GFC experience may motivate (re)insurers to stay with their traditional rules and time-consuming

procedures and to enhance them where necessary to deal even better with uncertainty and wrong incentives. Potentially useful principles are mainly assembled in the “Workarounds” subsections of the past three sections, let us just repeat a few keywords:

- minimum risk retention (alignment of interest)
- short risk transfer chains
- think in 50-year periods
- be sceptical when surrounded by a very optimistic crowd
- insurable interest.

Finally, this paper wants to focus once more on the situation of the individual practitioners in the financial industry – quants, management, and others. Most of us want to do a good job and at the same time would like to be comfortable at work, despite the hard decisions we have to make. What can we learn from the crisis?

- We probably must develop a better feeling for randomness. Although we have learned how modern statistics deal with extremes and dependencies, it could be that our intuition is still somewhat stuck in past times when the so-called “normal” distribution was considered the “normal” case and fluctuations could always be “diversified away” thanks to (quite) independent risks.
- When having to work with scarce data, we should insist on getting sufficient time to quantify the impact of the critical assumptions we have to make. We should preferably work with assumptions that can be tested against reality over time.
- We need better data and should make the effort to collect, structure, and maintain potentially useful information.
- We must learn from history. Although the world around us is in continuous change, there are still some useful lessons learned by our predecessors in past decades (or centuries). It would be a pity if we repeated those errors over and over again.
- Rather than creating illusions about manageability, we should acknowledge that the modern world is far more uncertain than we used to, and would still like to, believe. Maybe in some moments we are not able to handle more than 10% of the overall uncertainty and are nevertheless doing a good job – like a sailor who successfully steers a small vessel through a storm.

And we should learn to live happily in spite of all these uncertainties around us.

7. Epilogue

Apart from giving us new insights into pricing, underwriting, and risk management, the GFC could also change our view of the traditional investment policies of financial service providers like life insurers and pension funds. Typically, such long-term investors are bound to what is seen as conservative investing. In particular, they are often restricted to investment grade bonds (Caballero et al., 2017). As discussed in Section 2, in terms of cash flow, such bonds are analogous to very high (re)insurance layers, while low layers correspond to junk bonds. It is remarkable that in insurance high layers are traditionally felt to be the riskiest ones (“unbalanced”), while in finance the corresponding investment grade bonds are perceived as the safest option (almost “negligible” probability of default). In contrast, investors are concerned about junk bonds, while insurers don’t see low layers as fundamentally problematic.

Are the insurance and the capital market so different that such contradicting views can be adequate in their respective area? Or could there be circumstances when it would be good for investors to adopt the insurers’ view (or vice versa), at least when investors stay with long-term

investments (buy and hold), which is most similar to the insurers' way to assume risks? More to the point:

- If we have a (quite) efficient market and (quite) correct ratings and prices, does it really matter whether you invest in AAA or in junk, provided you buy the duration you need?
- If not, which market segment bears the higher risk of error and/or change? AAA or junk?
- Which market segment bears the higher systemic (or accumulation) risk? AAA or junk?

References

- Akseli, O. (2013). Securitisation, the financial crisis and the need for effective risk retention. *European Business Organization Law Review (EBOR)*, *14*(1), 1–27.
- Albrecher, H., Beirlant, J., & Teugels, J.L. (2017). *Reinsurance: Actuarial and Statistical Aspects*. Wiley.
- Boddy, C.R. (2011). The corporate psychopaths theory of the global financial crisis. *Journal of Business Ethics*, *102*(2), 255–259.
- Brazauskas, V., & Kleefeld, A. (2009). Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insurance: Mathematics and Economics*, *45*(3), 424–435.
- Caballero, R.J., Farhi, E., & Gourinchas, P.-O. (2017). The safe assets shortage conundrum. *Journal of Economic Perspectives*, *31*(3), 29–46.
- Donnelly, C., & Embrechts, P. (2010). The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *Astin Bulletin*, *40*(1), 1–33.
- Dowd, K., Cotter, J., Humphrey, C., & Woods, M. (2008). How unlucky is 25-sigma? *The Journal of Portfolio Management*, *34*(4), 76–80.
- Duffie, D. (2008). Innovations in credit risk transfer: implications for financial stability. *BIS Working Paper, No. 255*. Bank For International Settlements (BIS).
- EBA (2014). EBA report on securitisation risk retention, due diligence and disclosure. *European Banking Authority*, December 22.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling Extremal Events for Insurance and Finance*. Springer.
- Fackler, M. (2013). Reinventing Pareto: fits for both small and large losses. In *ASTIN Colloquium*.
- Frankland, R., Eshun, S., Hewitt, L., Jakhria, P., Jarvis, S., Rowe, A., Smith, A., Sharp, A., Sharpe, J., & Wilkins, T. (2014). Difficult risks and capital models. A report from the Extreme Events Working Party. *British Actuarial Journal*, *19*(3), 556–616.
- Gorton, G. (2009). The subprime panic. *European Financial Management*, *15*(1), 10–46.
- Haddrill, S., McLaren, M., Hudson, D., Ruddle, A., Patel, C., & Cornelius, M. (2016). *JFAR Review: Group Think*. IfoA Regulation Board on behalf of the Joint Forum on Actuarial Regulation. Institute and Faculty of Actuaries.
- Hellwig, M. F. (2009). Systemic risk in the financial sector: an analysis of the subprime-mortgage financial crisis. *De Economist*, *157*(2), 129–207.
- Jakhria, P. (2014). Difficult risks and capital models. A report from the Extreme Events Working Party. Abstract of the London Discussion. *British Actuarial Journal*, *19*(3), 617–635.
- Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: a cognitive perspective on risk taking. *Management Science*, *39*(1), 17–31.
- Klugman, S. A., Panjer, H.H., & Willmot, G.E. (2008). *Loss Models: from Data to Decisions*. Wiley.
- Krahn, J. P., & Wilde, C. (2017). Skin-in-the-game in ABS transactions: a critical review of policy options. *SAFE Policy White Paper No. 46*. Leibniz-Institut für Finanzmarktforschung SAFE.
- MacKenzie, D., & Spears, T. (2014a). 'A device for being able to book P&L': the organizational embedding of the Gaussian copula. *Social Studies of Science*, *44*(3), 418–440.
- MacKenzie, D., & Spears, T. (2014b). 'The formula that killed Wall Street': the Gaussian copula and modelling practices in investment banking. *Social Studies of Science*, *44*(3), 393–417.
- McNeil, A.J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
- O'Neill, W., Sharma, N., Carolan, M., & Charchafchi, Z. (2009). Coping with the CDS crisis: lessons learned from the LMX spiral. *Journal of Reinsurance*, *16*(2), 1–34.
- Pollock, A. J. (2018). *Finance and Philosophy: Why We're Always Surprised*. Paul Dry Books.
- Posner, E. A., & Weyl, E. G. (2013). An FDA for financial innovation: applying the insurable interest doctrine to twenty-first-century financial markets. *Northwestern University Law Review*, *107*(3), 1307–1357.

Puccetti, G. and Scherer, M. (2018). Copulas, credit portfolios, and the broken heart syndrome. An interview with David X. Li. *Dependence Modeling*, **6**(1):114–130.

Schmutz, M., & Doerr, R. R. (1998). *The Pareto Model in Property Reinsurance: Formulas and Applications*. Swiss Reinsurance Company.

Schwepcke, A., editor (2004). *Reinsurance: Principles and State of the Art – A Guidebook for Home Learners*. Verlag Versicherungswirtschaft.

Sinn, H.-W. (2010). *Casino Capitalism: how the Financial Crisis Came About and What Needs to be Done Now*. Oxford University Press.

Swiss Re (2002). *An Introduction to Reinsurance*. Swiss Reinsurance Company.

Tredger, E. R., Lo, J., Haria, S., Lau, H., Bonello, N., Hlavka, B., & Scullion, C. (2016). Bias, guess and expert judgement in actuarial work. *British Actuarial Journal*, **21**(3), 545–578.

Tsanakas, A., Beck, M. B., & Thompson, M. (2016). Taming uncertainty: the limits to quantification. *ASTIN Bulletin*, **46**(1), 1–7.

Weick, M., Hopthorow, T., Abrams, D., & Taylor-Gooby, P. (2012). Cognition: minding risks. *Lloyd's Emerging Risk Reports*.

Whitehouse, M. (2005). Slices of risk: how a formula ignited market that burned some big investors. *The Wall Street Journal*, September 12.

A. Mathematical Appendix

A.1. Diversification by pooling

Charts displayed in Figure 1 (Section 2): pools of n independent risks (or equivalently 1 risk in n years):

- $n = 10; 100; 1,000$.

Model for each individual risk:

- loss count distribution: Bernoulli.
- loss frequency 2%; loss size 1,000 (constant); expected loss 20; premium 25; average loss ratio 80%.

A.2. Risk with heavy severity tail

- ground-up frequency: 300.
- loss severity model: spliced exponential-Pareto (more exactly: $pExp-Par-0$ in the terminology of Fackler (2013)).

$$\bar{F}(x) = P(X > x) = \begin{cases} e^{-\frac{x}{\mu}}, & 0 \leq x < s \\ e^{-\frac{s}{\mu}} \left(\frac{s}{x}\right)^\alpha, & s \leq x \end{cases}$$

where $s = 1$ (e.g. in million Euros), $\mu = 0.145$, $\alpha = 1.5$.

Table A1 assembles some frequencies and key figures of the layers discussed in Section 3.4.

A.3. Retrocession XL for pool of layers

Setting:

- Suppose $2n$ copies of a reinsurance layer c xs d , where c is the limit per loss and per year, having loss count N , loss frequency $\lambda := E(N)$, aggregate loss S , and risk premium $E(S)$.
- Note that the trigger variable $T := \chi_{\{S > 0\}} = \chi_{\{N > 0\}}$ is Bernoulli distributed with parameter $p := E(T) = P(N > 0) \leq \lambda$.
- Suppose the pool's loss $\sum_{i=1}^{2n} S_i$ (having maximum $2nc$) is protected by a retrocession layer nc xs nc , which pays:

$$Z := \min \left(\left(\sum_{i=1}^{2n} S_i - nc \right)^+, nc \right)$$

Table A1. Frequencies and pure premiums of layers

Threshold	Frequency [%]
1	30.00
2	10.62
3	5.77
4	3.75
5	2.68
6	2.04

Layer	Frequency [%]	Average severity	Expected loss	Share [%]
2 xs 2	10.62	1.17	0.124	31
2 xs 4	3.75	1.47	0.055	69
4 xs 2	10.62	1.69	0.179	100

We give an upper bound for the risk premium of the retro layer in the case that the S_i are independent. Key idea: to hit the retro layer, more than half of the underlying reinsurance layers must suffer a loss. Due to the independence we have $\sum_{i=1}^{2n} T_i \sim \text{Binomial}(2n, p)$, which yields:

$$\begin{aligned}
 P(Z > 0) &= P\left(\sum_{i=1}^{2n} S_i > nc\right) \leq P\left(\sum_{i=1}^{2n} T_i > n\right) = \sum_{k=n+1}^{2n} P\left(\sum_{i=1}^{2n} T_i = k\right) = \sum_{k=n+1}^{2n} \binom{2n}{k} p^k (1-p)^{2n-k} \\
 &< p^{n+1} \sum_{k=n+1}^{2n} \binom{2n}{k} < \frac{p^{n+1}}{2} \sum_{k=0}^{2n} \binom{2n}{k} = \frac{p^{n+1}}{2} 2^{2n} = \frac{p}{2} (4p)^n
 \end{aligned}$$

So, for the risk premium of the retro layer, we have:

$$E(Z) \leq nc P(Z > 0) < \frac{npc}{2} (4p)^n$$

If we relate this to the aggregated risk premium of the underlying pool of layers, we get:

$$\frac{E(Z)}{2nE(S)} < \frac{pc}{4E(S)} (4p)^n \leq \frac{\lambda c}{4E(S)} (4\lambda)^n$$

Let us calculate the last upper bound for the example in Section 3.4. We have $n = 50$ and $c = 4$. The layer loss S results from the model presented in Section A.2; from Table A1, we take $E(S) = 0.179$ and $\lambda = 10.62\%$. Plugging in we get:

$$\frac{E(Z)}{2nE(S)} < 1.42 \cdot 10^{-19}$$

A.4. Inferring VaR from 10 observations via Maximum Likelihood

Simple case, having only one tail parameter to estimate:

- spliced severity model: small/medium vs large losses.
- tail model above $s = 1$ (e.g. in million Euros): exponential or Pareto.
- model below s unspecified (has no impact on VaR).

Table A2. Empirical loss record and statistics

Loss	x_i	y_i	z_i
1	0.5		
2	1.7	0.7	0.53
3	3.4	2.4	1.22
4	2.2	1.2	0.79
5	0.9		
6	1.2	0.2	0.18
7	1.9	0.9	0.64
8	2.8	1.8	1.03
9	0.3		
10a	12.3	11.3	2.51
10b	12.4	11.4	2.52

Scenario	a	b
\bar{y}	2.64	2.66
$\hat{\mu}$	2.64	2.66
$NLL(\hat{\mu})$	13.80	13.84
$VaR(\hat{\mu})$	12.2	12.3
\bar{z}	0.99	0.99
$\hat{\alpha}$	1.01	1.01
$NLL(\hat{\alpha})$	13.81	13.83
$VaR(\hat{\alpha})$	66.1	66.5
Best fit	Exp	Par
\widehat{VaR}	12.2	66.5

$$\bar{F}(x) = P(X > x) = \begin{cases} \dots, & 0 \leq x < s, \\ r \bar{F}(x|X > s), & s \leq x, \end{cases} \quad \bar{F}(x|X > s) = \begin{cases} e^{-\frac{x-s}{\mu}}, & \text{Exponential} \\ (\frac{x}{s})^\alpha, & \text{Pareto} \end{cases}$$

Suppose we observe n losses x_i , out of which k exceed s (large losses). For the latter, we define $y_i := x_i - s$, $z_i := \ln(x_i/s)$, which will be used for tail inference. The estimate for the tail weight is $\hat{r} = k/n$.

Exponential: If the x_i have an exponential tail, the y_i are exponentially distributed and for their average \bar{y} we have $\hat{\mu} = \bar{y}$. For the negative log-likelihood, one gets easily that:

$$NLL(\mu) = \frac{1}{\mu} \sum_{i: x_i > s} y_i + k \ln(\mu), \quad NLL(\hat{\mu}) = k \{1 + \ln(\bar{y})\}$$

Pareto: If the x_i have a Pareto tail, the z_i are exponentially distributed and for their average \bar{z} we have $\hat{\alpha} = 1/\bar{z}$. For the negative log-likelihood, one gets easily that:

$$NLL(\alpha) = (1 + \alpha) \sum_{i: x_i > s} \ln(x_i) - k \ln(\alpha) - k\alpha \ln(s) = k\{(1 + \alpha)\bar{z} + \ln(s) - \ln(\alpha)\}$$

$$NLL(\hat{\alpha}) = k\{1 + \bar{z} + \ln(\bar{z}) + \ln(s)\}$$

The key figures of an example with 10 losses are assembled in Table A2. We see that a minimal change of the largest loss (10a/b) alters the preferred model, which leads to a huge change of the estimated $VaR = \bar{F}^{-1}(1\%)$.