

Big Data is coming – are you ready?
Insurance principle and actuaries in the age of fintech

Insurance is based on the idea of pooling of risks: everyone in the insured pool pays a modest premium so that the unlucky ones in the pool encountering a loss will be compensated. Risks accepted into the pool must be uncertain. It must be possible to quantify their extent and probability. Providing that the risks are (at least to a large extent) uncorrelated the law of large numbers makes the pool manageable – the larger the pool the less the losses vary around the average. An additional concern in the management of the pool is to take care of moral hazard – members of the pool should take care to manage their risks responsibly. Especially they should not be able to benefit from the pool at the expense of other members through their own actions (or inactions).

Through all the history of insurance insurers have used all possible tools to manage insured pools as well as possible. Initially this has meant that insured risks are analysed and differentiated into separate categories based on their riskiness. This has made it possible to set an as fair as possible premium to each risk. Insurance in its voluntary form does not contain inappropriate cross subsidies from one member of the pool to another member (It needs to be noted that in compulsory or social insurance such cross subsidies are possible – and also when the use of gender as a tariff factor became forbidden in the EU also voluntary insurance has more cross subsidies than before).

The risks in the pool are monitored even when they are already in the pool. This plays a role when insurers calculate their technical provisions that show how much assets they need to have in order to cover their risk position. In the extreme a member of the pool can become too risky and will be excluded from the pool (although legislation often has strict limitations for this exclusion). It has been said in some occasions that the emergence of fintech and especially the phenomenon of Big Data with developing data analytics will make insurance obsolete. This is based on the idea that with new tools it will become possible to forecast the future so minutely that there is no room for uncertainty and no role for insurance. While it is true that many new challenges to the insurance principle are emerging with Big Data we feel the insurance principle is alive and well for the foreseeable future. Some key issues are discussed in the sequel.

Big Data denotes the idea that the exponentially increasing amount of digitally stored data creates a radically new world. In addition to the volume of data the concept includes the advances in the following areas:

- veracity (meaning data quality needs to be checked and biases, distortions and noise should be cleaned),
- variability (data comes from many sources and is often less structured), and
- velocity (due to Moore's law it is possible to use increasingly sophisticated methods to analyse data, with the cost going down).

It can already now be said that the possibilities involved in this development present insurers revolutionary new tools to manage insured pools. Traditionally actuaries have used simple proxies (like age, gender, address, smoker/non-smoker) to differentiate risks into different tariff classes. If

however the insurer would today know a person's shopping habits for ten years, what this person buys from the pharmacy, how often this person goes to the gym etc., one could (at least in theory) say it is unnecessary to know the age or gender of this person. It needs to be said that today's analytical knowledge is catastrophically lacking in this area: there is a huge lack of understanding the causality of certain data leading to different outcomes.

It needs to be added that the revolution is not only about the volume of data but also very much of velocity: in earlier times even when there was data exact methods would have used too much computing power, leading to the use of inexact/proxy methods. Also in earlier times data was a very expensive resource meaning it made a lot of sense to create simpler proxies.

We'll start this analysis from the idea of a "perfect" world. In an ideal world we would have complete information and our analytical tools would make perfect predictions of the future. Would we be able to say exactly when a house burns, when two vehicles crash or when a person contracts a disease or dies? No, we would still only have probabilities, albeit more exact ones than today, of what will happen. We would still not be omniscient and omnipotent gods knowing in advance how things turn out.

We would certainly be able to know the probabilities better. We would be able to differentiate insured risks into more exact categories. We would be able to set fairer premiums for the risks. But we would still need to pool the risks to benefit from the stabilising effect of the law of large numbers. The insurance principle would be a stronger, not a weaker tool to manage our destinies.

One of the main problems in utilising the possibilities is our lack of proficiency in data analytics. One of the main dangers and challenges is our limited understanding of different causalities. From huge amounts of data we can find many interesting correlations between different things. Correlation means two things happening more or less together. But correlation does not mean one can deduce an outcome. To forecast one needs causality, i.e. the knowledge of something leading to another. It is extremely dangerous to misinterpret correlations as causalities. Without advances in data analytics we cannot make good use of the new possibilities.

Insurance is always based on modelling the reality (model can here also be called an algorithm). These models are always imperfect. With increasing amounts of data it will be increasingly difficult to make certain that the models created will reflect the reality correctly. The danger – connected heavily to what is said on correlation and causality – is that the models might only be adapted to the available data with no predictive power to situations not included in this data. Big Data will make increasingly complicated models possible with an increasing need to make certain the models have some real sense.

Traditional understanding in insurance is that the information asymmetry benefits the customer. With this we mean that the customer always knows more of his/her situation compared to what the insurer knows. This situation might be overturned with the emergence of big data – the insurer with huge amounts of data and advanced analytical capabilities might be in a fundamentally better position vis-à-vis the customer. This is a threat if this leads to insurers misusing this situation to the detriment of the customer. In the extreme insurers will be able to segment risks into many more categories than what is possible today, with some risks becoming uninsurable. It is unclear how much regulatory attention this might need in the future.

As a conclusion we feel the insurance principle will retain its applicability also with the emergence of Big Data. Big Data will create possibilities and challenges to its application but generally there are more possibilities than threats. In the increasingly complicated and interrelated world of tomorrow there will be an increased need for pooling of risks and better techniques available to make this pooling possible.

The emergence of Big Data creates a radically new landscape. Such paradigm shifts have occurred earlier. Laplace presented scientific advances with the famous quote "*Je n'avais pas besoin de cette hypothèse-là*" (referring to God). During the time of the Soviet Union some Russian scientists said that stochastics is only a science needed before full communism where all things will be certain. From the 1980's, advances in genetics and sequencing of the human genome have generated thoughts of full certainty of human life. The insurance principle has easily survived these challenges and will do so in the age of Big Data.