OXFORD MARTIN SCHOOL

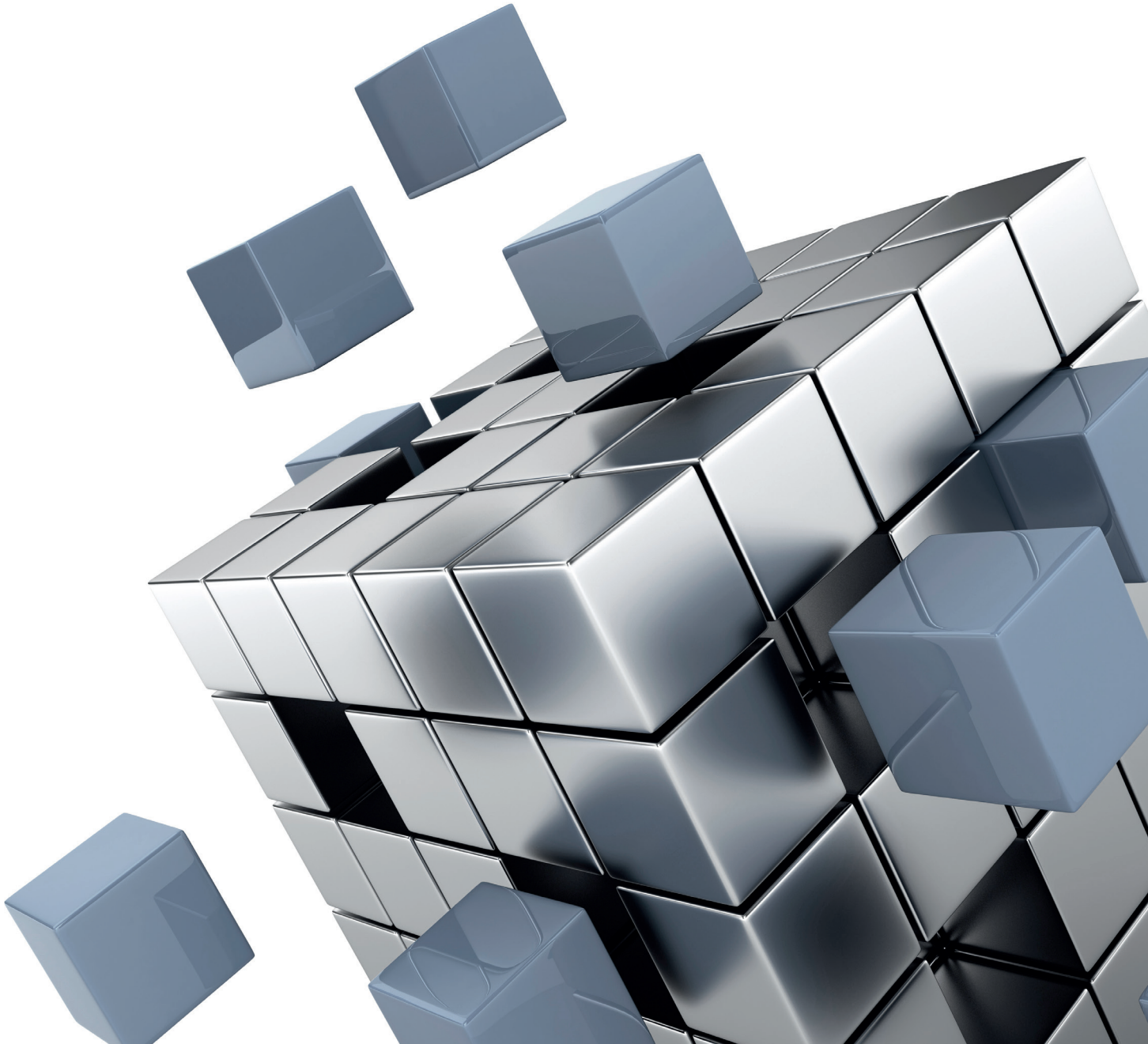
UNIVERSITY OF OXFORD

Amlin

Systemic Risk of Modelling in Insurance

Did your model tell you all models are wrong?

White paper by the Systemic Risk of Modelling Working Party



Contents

If you are interested in finding out more about Systemic Risk of Modelling, please email: SRM@amlin.com
The Leadenhall Building, 122 Leadenhall Street, London EC3V 4AG
Switchboard +44(0) 207 746 1000

Foreword by Simon Beale

In his book The Butterfly Defect, Professor Ian Goldin observes that "Systemic risk cannot be removed because it is endemic to globalisation. It is a process to be managed, and not a problem to be solved."
It is primarily for this reason that Amlin and the Oxford Martin School (OMS) have worked with industry experts to develop a practical and applied method to consider ways of encouraging the quantification, monitoring, reporting and management of the systemic risk of modelling.

Systemic risk is increasing in relevance due to the growing level of globalisation, interconnectedness, and speed of business transactions in our world. Whilst these trends are beneficial they also introduce new complexities to some existing risks. The natural response to a more risky and interconnected world is to try quantify it and thus modelling as a means to better understand risk is becoming increasingly common and in itself, complex. Systemic risk of modelling (SRoM) is by its very nature often difficult to quantify partly due to the way in which many, often sophisticated, systems interact with each other.

This paper is specifically focused on a practical solution for our industry, with the objective to design a risk scorecard for the SRoM - a practical way to measure the amount of systemic risk introduced from modelling leveraging the Amlin - OMS collaborative research and findings to date. The SRoM scorecard has been designed less towards an exact risk measure and more towards providing an indication of whether certain actions and practices are aligned with reducing systemic risk of modelling.

Our announcement to introduce and address systemic risk using an applied and practical method with leading figures from the insurance market was communicated at our London event last year. I am pleased to convey that the strong interest was self-evident as experts from both academic research and the insurance industry volunteered to contribute and co-author this paper. As a result of this concerted research we are better positioned to design a SRoM scorecard that firms and regulators can use to monitor and manage systemic sources and measures of systemic risk. In addition practical solutions towards measuring whether risk management decisions and modelling practices are aligned with reducing or heightening systemic risk were developed and are incorporated in this paper.

Amlin and the Oxford Martin School have worked collaboratively with this group of experts and market practitioners with representation from life and general insurers, reinsurers, catastrophe model vendors, brokers, regulators and consultants. By bringing together the 'best minds' in the business with the 'best minds' in academia an effective and meaningful SRoM scorecard has been developed which can be used by firms and regulators as part of their toolkit to better monitor and manage risk.

The technical expertise and support of the individuals involved has been invaluable and I am grateful to all those who have helped to share ideas and best practices for this paper. I believe the SRoM scorecard is another significant step in managing the systemic risks we face and ensuring that our reliance on the increasing amalgamation of models in the insurance industry is managed more effectively.

Simon Beale, Chief Underwriting Officer, Amlin plc.

Foreword by Ian Goldin

As we move into the 21st century the world is becoming more connected, more complex and more uncertain. The latest wave of globalization has integrated markets and finance while the information revolution has compressed time and space. The result is that both virtual and physical proximity has increased, in general to great benefit. Similarly the network structures underlying society and technology have grown to become more complex, interdependent and integrated than ever before, facilitating the transmission of material, capital and knowledge at a high degree of efficiency.

Yet this connectedness and complexity increases uncertainty: more factors from more distant places can influence events, crises can unfold far faster than before, and the sheer flood of information taxes the ability to distinguish the signal from the noise in a timely manner. The result of these trends is a growth of systemic risk that challenges the benefits that globalisation has produced.

As the financial crisis demonstrated, there is a real risk that governance can fall behind the growth of new models and instruments: tools that may reduce risk in certain circumstances may have unexpected and destructive ramifications when used unwisely. Making strong assumptions and simplifications in models that failed to capture the complexity and systemic nature of finance produced a form of collective overconfidence that led to crisis. In order to reduce systemic risk, risk governance needs to be strengthened. There is a need for new models that take account of the integration and complexity of network structures, as well as for transparent communications about choices, risks, and uncertainty of policy alternatives, improved risk measurement, and promotion of resiliency and sustainability.

In contributing to improvements in risk management the Oxford Martin School has collaborated with the Industry Working Party on Systemic Risk of Risk Modelling to develop this report. The aim has been to understand how different factors build up systemic risk when risk models are used, and approach a way of detecting the risk. Understanding how the fallibility of human thinking, competitive pressures, inadequate governance and weak modelling processes interact to produce systemic risk is a first step towards measuring and mitigating it. It is my hope that this document will in the long run help improve systemic risk governance not only in insurance but in other institutions where risk models are used.

Professor Ian Goldin, Director of the Oxford Martin School at the University of Oxford.

Author of The Butterfly Defect, how globalisation creates systemic risks, and what to do about it.

Contributors

The views expressed in this paper were held under the Chatham House Rule and therefore do not necessarily reflect the views of the organisations to which the working party members belong.

Academics		
Owain Evans	Research Fellow	Oxford Martin School
Ben Levinstein	Research Fellow	Oxford Martin School
Anders Sandberg	Senior Research Fellow	Oxford Martin School
Andrew Snyder-Beattie	FHI Research Director	Oxford Martin School
Cecilia Tilli	Programme Manager	Oxford Martin School
Feng Zhou	Research Fellow	Oxford Martin School
Brokers		
Jayant Khadilkar	Partner	Tiger Risk Partners
Vladimir Kostadinov	Associate	Tiger Risk Partners
Heather Roscoe	Associate	Tiger Risk Partners
David Simmons	Managing Director Analytics	Willis
Consultancy		
Domenico del Re	Director	PwC
General Insurance		
Mark Christensen	Head of Catastrophe Management	ACE European Group
JB Crozet	Head of Underwriting Modelling	Amlin plc
Alan Godfrey	Head of Exposure Management	ASTA Managing Agency Ltd
Matt Harrison	Syndicate Exposure Manager	Hiscox
Andrea Hughes	Delivery Manager, Underwriting Modelling	Amlin plc
Andrew Leach	Head of Catastrophe Modelling - Europe	Travelers Insurance
Alan Milroy	International Property Catastrophe Underwriter	XL Catlin
Catherine Pigott	Natural Perils Actuary	XL Catlin
David Singh	Underwriting Exposure Manager	Amlin plc
Gemma Smyth	Exposure Manager	Charles Taylor Managing Agency Ltd
Stav Tsielepis	Vice President, Risk Management Actuary	Arch
Model Vendors		
Gabriela Chavez-Lopez	Account Director	CoreLogic
Andrew Coburn	Senior Vice President	Risk Management Solutions
Shane Latchman	Senior Manager, Research and Client Services Group	AIR Worldwide
Milan Simic	Senior Vice President and Managing Director	AIR Worldwide
Mohammed Zolfaghari	Director	Cat Risk Solutions Ltd
Life Insurance		
David Kendix	Head of Insurance Risk & Model Oversight	Prudential plc
Regulator		
Dimitris Papachristou	Chief Actuary (Research) General Insurance	Bank of England
Rating Agency		
Miroslav Petkov	Director	Standard and Poor's Ratings Services

Introduction

1.1 Purpose

The purpose of this White Paper is to help the readers understand the systemic risk associated with Modelling practices within the insurance industry.

The term "Systemic Risk" is often used when describing the Global Financial Crisis of 2007-2008. The catalyst for this financial meltdown was the bursting of the U.S. housing bubble which peaked in 2004, causing the value of securities tied to the U.S. housing market to plummet and damage financial institutions on a global scale.
Core to the magnitude of the losses experienced by these events was the Modelling assumptions upon which these securities were valued, the extent to which these assumptions were applied across markets and the sensitivity of these assets' values to those assumptions – assumptions that turned out to be unreliable. These Modelling shortcomings took a Systemic nature because the whole mortgage-backed industry was using more or less the same models, as they had been "institutionalised" by credit rating agencies. When the model assumptions began to fail the economic effects further amplified the model failure and losses.

This paper is specifically focused on the practical understanding of Systemic Risk of Modelling and development of practical solutions for managing such risk within the insurance industry. Attempts to understand and manage Systemic Risk can be considered to fall into 4 categories.

1. **Existence of risk:** this is the process whereby we acknowledge, quantify and qualify the elements and drivers of systemic risk which occur or could potentially occur within the insurance industries modelling practices.

2. **Observation of risk:** the process whereby Systemic Risk factors can be observed and tracked over time, to ensure that the process of Systemic Risk Management is a proactive rather than reactive mechanism.

3. **Risk triggers:** the insurance industry is founded on the basis of taking on board risk and the existence of risk is core to its operation. Risk triggers however act as a mechanism for indicating where risk levels may be approaching or exceeding agreed risk tolerance levels.

4. **Mitigation of risk:** Certain risk factors may be able to be diminished at source or managed by underwriting or operational practices or processes. Where this is not the case, it may be possible to minimise the potential effects of such risk via arbitrage

This White Paper aims to provide the reader with:
- A 'framework' to understand the Systemic Risk of Modelling and build a Risk Management process for it; and

- The design of a 'Risk Scorecard for the Systemic Risk of Modelling' as a practical way to measure and raise awareness about Systemic Risk of Modelling within organisations.

Our target audience is practitioners in the insurance industry – whether underwriters, brokers, modellers or executives - but also regulators and people in other organisations that use risk models.
We are hoping that this White Paper will contribute towards developing guidelines for sustainable Modelling practices within organisations and across the industry, in particular

- How do we manage this risk, which is behavioural and operational in nature?

- How do we educate users of model results about the potential pitfalls?

- How do we develop a sustainable, robust usage of models within the insurance industry?

1.2 Scope

The context of this report is general risk modelling in insurance. It aims at being independent of the type of insurance: model risk exists in all domains, whether data-rich life, casualty and commodity insurance models, the more data-poor catastrophe models, capital models, or other actuarial models. However, there may possibly be wider applications: we are conscious about the growing importance of risk models in civil defence, planning, governance and forecasting. Many of the risk factors are the same, and the experience of the insurance world can be helpful to avoid the pitfalls of modelling while reaping the benefits.

1.3 The Authors

This White Paper was drafted by the Systemic Risk of Modelling Working Party, as a collaboration forum composed of academic researchers and industry practitioners.

The idea behind forming the Working Party was to bring together the 'best minds' in the business with the 'best minds' in academia, in order to tackle the complex and significant issue of Systemic Risk of Modelling.

Whilst the Working Party has representation across many parts of our industry (including general insurance, life insurance, brokers, model vendor, regulator, rating agency, actuarial consultancy), there is a bias in experience amongst its members towards London Market Catastrophe Modelling.

Contributions have been made based on the Chatham House rule and do not reflect the views of the participating organisations.

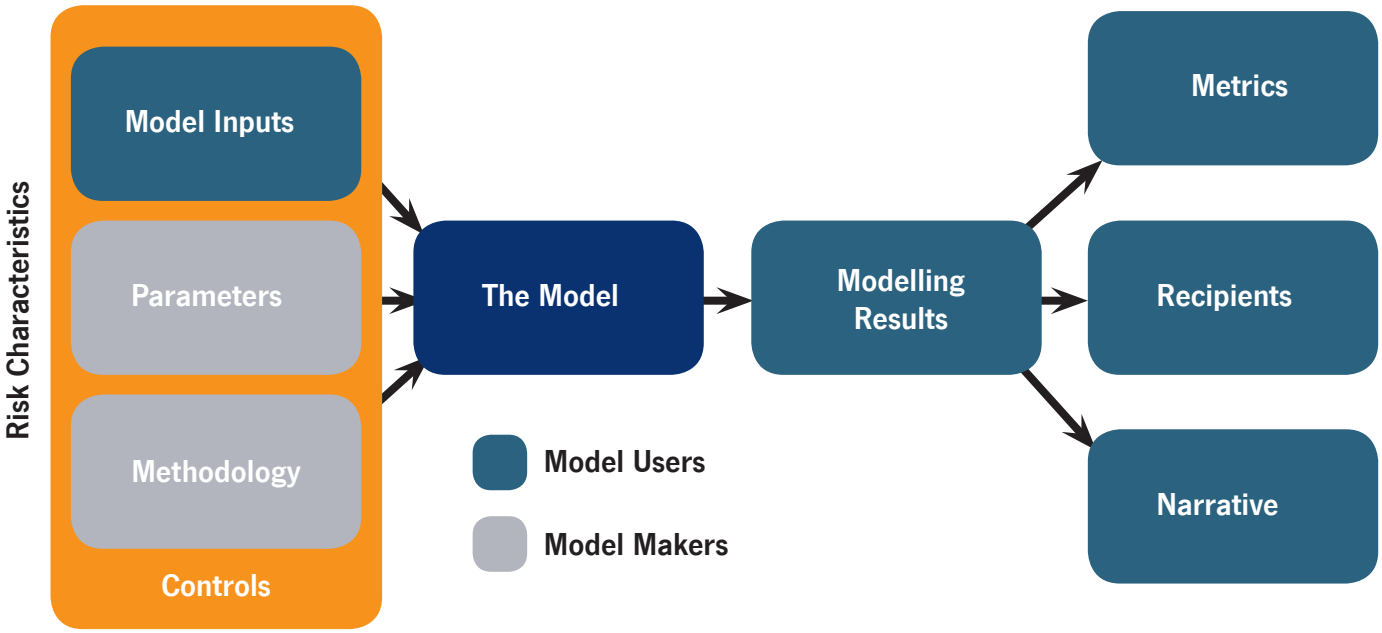
1.4 What is the Systemic Risk of Modelling?

The Modelling Process

A model is a representation of a selected part of the world, often based on a theory of how it works (or appears to behave).

The Modelling Process is the activity of building a model to fit the world, but also the use this model to learn and predict the part that is being modelled.

Models are part of larger processes of decision making, organisation and explanation. The actual mathematics and software is not the only relevant part: how they are being used can matter far more for the outcomes we care about.



The Modelling Process for a risk model has several components:

Component	Definition	Example 1: General Insurance Catastrophe Mode	Example 2: Life Assurance Internal Mode
Risk Characteristics	Known or assumed properties of the risk being modelled.	Apartment on the 25th floor of a tower block.	Policy holder aged 50 withknown medical problems
Model Inputs	Information on the risk used by the model.	Street address, sums insured (buildings, contents and BI), deductibles.	Demographic information (age, gender), health information, location, contract type, deductibles.
Parameters	Elements of the model which define results for a given set of inputs.	Probability distribution of hazard rate, intensity and location, damage functions, correlations between perils.	Mortality risk as a function of class, distribution and correlation structure of external factors.
Methodology	The theoretical approach and assumptions combined to build a model, and how the model derives the results from the inputs.	Stochastic (Monte Carlo) simulation models.	Actuarial models based on mortality tables.
Controls	Methods to plan and validate the model creation and usage.	Dependent validation, peer reviews and independent model validation, audits, peer reviews and back testing.	Dependent validation, peer reviews and independent model validation, audits, peer reviews and back testing.
Modelling Results	Estimates of the risk/price probability distribution as a function of the model inputs.	Average annual losses, Exceedance Probability Curves.	Profit and loss distribution, Solvency Capital Requirement.
Metrics	Measures of the key modelled output used for risk management and monitoring.	US Windstorm 1 in 200 VaR, 1 in 100 TVaR.	Profit and Loss 1 in 200 VaR.
Recipients	Stakeholders that make use of the modelling results directly or indirectly.	Underwriters, brokers, actuaries and regulators.	Underwriters, brokers, actuaries and regulators.
Model Users	People responsible for operating the Model, feeding Model inputs and extracting/analysing Modelling Results.	Catastrophe Modellers within an insurance organisation.	Actuaries running the Internal Model.
Model Makers	People responsible for the Methodology and Parameters in the Model.	Vendors of Catastrophe Models, Natural Hazard Experts.	Actuaries specialised in building the Internal Model.
Narrative	Subject matter expert review, commentary and analysis.	Board packs, model change commentary.	ORSA, qualitative change analysis.

We will use this “canonical” Modelling Process throughout the White Paper.

Benefits of Modelling

The benefits of quantitative models are undeniable, which explains the speed of their adoption by the market and its regulators. By providing a consistent, informed assessment of the risks within our business, quantitative models have helped (re)insurers on several fronts:

- Risk Management: the ability to manage risk on a probabilistic basis, and compare the riskiness of very different types of exposures (e.g. life assurance vs. property catastrophe);

- Portfolio Management: the ability to measure the risk-return profile of the current portfolio, and produce alternative "what if" scenarios;

- Technical Pricing: the ability to measure the expected cost associated with a specific contract, and compare the relative value of alternative policy structures.

These benefits have helped to partially "de-risk" the business of (re)insurance, by providing a control framework thereby lowering our "cost of capital" and enabling more affordable (re)insurance in the market.

Systemic Risk

Systemic Risk can be defined in many ways, but a useful rough definition is "risk that happens in a system because of the way its parts interact, rather than faults in the parts themselves." The system is vulnerable to self-reinforcing joint risks that can spread from part to part, affecting the function of the entire system, often with massive real-world consequences[1].

A key feature that emerges is that parts of a system that individually may function well become vulnerable to a joint risk when connected, leading to a spread in risk that potentially affects the entire system.

It provides a unique challenge as, unlike other risks, adaptation and risk mitigation (including regulation) are not separate from the system, and can actually increase the systemic risk. Additionally much of the risk comes from the structure of the system, which is often constrained, making strong changes infeasible.

[1] Anders Sandberg, Nick Beckstead, Stuart Armstrong. Defining systemic risk, In Systemic Risk of Modelling, report from the FHI-Amlin Systemic Risk of Risk Modelling Collaboration. Oxford University. 2014

Examples of Systemic Risk

- Financial asset price bubbles: asset prices begin to increase at an accelerating pace, triggered by fundamental or financial innovation. Investors pile in to take advantage, further increasing the price.
  It seems rational for agents to join in; especially if they expect that they can pull out in time or receive bailouts. Speculation and overconfidence ensue until the price crashes, often impacting the economy outside the financial sector[2]. The fact that bubbles have occurred numerous times in the past – from the Dutch tulip bulb mania 1634-37 to the recent US Housing bubble – does not strongly deter new bubbles.

Tulipmania

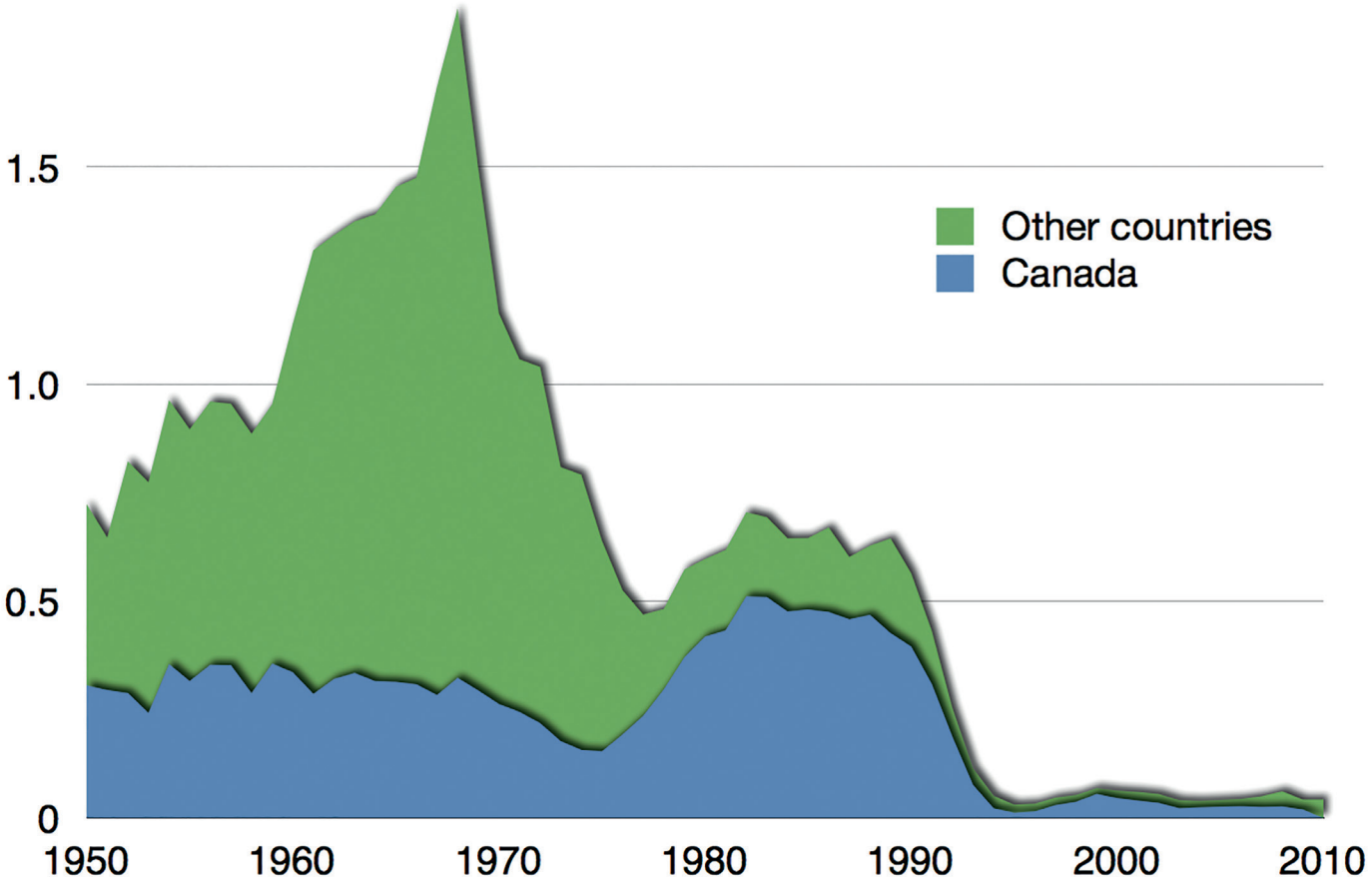
By the 1620s tulips had become fashionable in the Netherlands, with a high demand for bulbs from rare varieties but also uncertainty in their future value. One of the first futures market emerged, first for hedging, but soon speculators joined the gardeners. The tulip futures were less strongly regulated than normal investments or insurance and had few protections. Speculation in rare bulbs expanded and peaked in 1634-1637. At its peak, single bulbs could sell for the price of a building. In February 1637 tulip prices collapsed, and in April the authorities suspended all the futures contracts. While the severity of the event has been contested it was one of the first documented financial bubbles. The bubble was partially caused by normal speculation – people hoping to make a windfall and crowding in when they saw their neighbours' apparently getting rich – but also through inexperience with the new financial instruments.*

*The classic account of Tulipomania is Mackay, C. (1841), Chapter 3: Tulipomania. In Memoirs of Extraordinary Popular Delusions and the Madness of Crowds, London: Richard Bentle. For one modern economist's view, see Day, C. C. (2004). Is There a Tulip in Your Future?:

Ruminations on Tulip Mania and the Innovative Dutch Futures Markets. Journal des Economistes et des Etudes Humaines, 14(2), 151-170.

[2] Brunnermeier, M. K., & Oehmke, M. (2012). Bubbles, financial crises, and systemic risk (No. w18398). National Bureau of Economic Research. models (for example, by changing how risks are investigated and modelled, or how models are seen).

- Collapse of the Atlantic northwest cod fishery: cod fishing around Newfoundland had historically been very profitable. It was in each individual fisher's best interest to catch as much as possible. From the 1950s onwards new fishing technology arrived that enabled more efficient trawling, boosting yields – but also depleting cod stocks and disrupting the essential ecosystem. Uncertainties of the underlying situation and strong vested interests made remedial action too slow, and in 1992 cod numbers fell to 1% of their previous levels. This low state has been relatively stable and has only slowly recovered.



Source: Time series for the collapse of the Atlantic northwest cod stock, capture in million tonnes with Canadian data presented separately. Based on data sourced from the FishStat database 2 May 2012. Author:Epipelagic

Systemic Risk of Modelling in Insurance

Most discussion of systemic risk in the financial sector tends to focus on how institutions become connected through economic ties and how this can cause contagion. The insurance industry is safer than most financial industries because of its nature[3]. However, there is one area that is often overlooked as a source of systemic risk: risk modelling itself.

A risk model intends to make probability estimates of a risk, which will then be used to make decisions. If it underestimates the probability of a risk, actions will be taken in false confidence. If it overestimates risk, resources will be misallocated. If it causes correlated actions across an organisation or market, systematical[4] or systemic risk emerges[5].

[3] Geneva Association. (2010). Systemic risk in insurance: an analysis of insurance and financial stability. Special Report of the Geneva Association Systemic Risk Working Group, March.
[4] Systematical risk is distinct from systemic risk. A systematical risk is a simultaneous shock to the whole market that causes adverse effects: the cause is external to the system rather than internal.
[5] There are also possible feedbacks. The model can cause actions that change the risk itself, making the model obsolete and inaccurate (for example, the Year 2000 problem triggered remedial actions that reduced the risk, while bond rating models in the CDO market contributed to a bubble that changed the underlying risk for the worse). It can also cause actions that affect future reliability of risk models (for example, by changing how risks are investigated and modelled, or how models are seen).

Model Sources of Systemic Risk of Modelling

The large investments required to build sophisticated representations of (re)insurers' risks and the scalability of quantitative models, point to significant economies of scale from centralising and outsourcing their development to third-party vendors.

For instance, many (re)insurers license proprietary Economic Scenario Generators or Catastrophe Models from third-party vendors who generate the investment in talent and R&D.

While the financial benefits of outsourcing model-building to third-party vendors are often clear, the associated "outsourcing of cognition" presents some challenges in itself:

- The divergence in principal-agent interests might lead third-party vendors to be influenced by other priorities than modelling quality (e.g. production costs, sales potential, social and political context etc.);
- (Re)insurers have reduced incentives to invest in modelling knowledge and talent, to the point that their decision-makers could become over-reliant on the "autopilot" and unable to critique or even function without it;
- The oligopolistic nature of markets with large economies of scale, allows the few players to be more authoritative as central source of knowledge, than justified by the quality of their models alone.

Unfortunately, the more the industry tends to rely on a single source of knowledge, the smaller the upside when it gets things right and the greater the downside when it gets things wrong (as, one day, it inevitably will).

Organisational Sources of Systemic Risk of Modelling

The Internal Models in the UK ICAS and the coming EU Solvency II regimes (mimicking the Basel II regulations for financial institutions, by introducing Internal Models for regulatory capital setting) have, however, taken a widely different stance. In essence, the setting of minimum capital requirements is outsourced to the (re)insurer if its Internal Model is approved by the regulator:

- Internal Model Approval requires the regulator to be comfortable with the model, which could limit the range of potential approaches and possibly introduce "asymmetrical error checking" (i.e. mostly scrutinising the models which do not fit expectations or preferences);
- The Documentation Standards require sufficient details to enable the Internal Model to be justifiable to a third party, possibly restricting reliance on those areas of expert judgment for which documentary evidence is sparse, thereby slowing down the adoption of innovative approaches; and
- The Use Test requires that the Internal Model informs risk management and key decision processes, which is likely to restrict reliance placed on outputs from alternative models within the organisation.

The risk for our industry is that we are unconsciously dis-incentivising the emergence of alternative approaches, which are vital for a fully functional evolutionary process.

The result is that, as an industry, we have become dependent on the same models used within organisations and across organisations. This means that we are putting all our eggs in the same basket when it comes to Modelling, and we are therefore exposing ourselves to "rare but extreme" model failures.

Behavioural Sources of Systemic Risk of Modelling

The existence of model error, popularised by George Box's famous quote "all models are wrong but some are useful", is reasonably understood by practitioners in the market.

Our industry is, however, much less familiar with the risks arising from the behavioural aspects of the modelling process; or, in other words: how, in human hands, "all models are wrong, but even the useful ones can be misused".

Quantitative models have the significant advantage of scaling up with technological progress. Unlike expert judgment which is limited in speed and footprint, models become faster and more advanced as technology improves. Often, however, the gains in calculation speed are translated into a higher reporting frequency without necessarily a full appreciation for the critical, qualitative difference between, for example:

- A Chief Underwriting Officer receiving a quarterly report on the risk profile of the portfolio, supported by qualitative commentary from his Chief Pricing Actuary highlighting the limitations of the analysis; and

- The same Chief Underwriting Officer accessing the same figures daily, on a self-service basis at the press of a button.

These two types of reporting have a purpose adapted to different tasks. To draw a parallel with Daniel Kahneman's "Thinking, Fast and Slow": the slower and more deliberative approach is better adapted to more strategic situations, while the faster, instinctive reporting is best suited to monitoring contexts.

Without the awareness of this distinction, the temptation is great, however, for the Chief Underwriting Officer to rely on the faster, automated reporting for strategic decisions; leading to a "dumbing down" of the decision making process as a result of technological automation. Unlike expert judgment, quantitative models are based on transparent assumptions, which can be adapted in order to improve predictive power or adapt to environmental changes over time. Similarly, a model identified to not be fit-for-purpose would quickly be disregarded if it did not adapt appropriately.

This evolutionary process is a powerful force, which has helped our industry get better and better models over time. But we must recognise that the institutionalisation of quantitative models can lead to structural groupthink and "limit the gene pool" by reducing the potential for model diversity.

Historically, regulatory frameworks have not interfered with (re)insurers' freedom to select and use models as they deemed fit. The regulatory rules for setting minimum capital requirements complemented the internal risk management perspective with an independent view and an additional layer of defence.

1.5 Structure of the White Paper

- **The Framework section** is an explanation of the key drivers of the Systemic Risk of Modelling, split into modelling, organisational and behavioural sources of risk. It provides the foundations for the design of the scorecard.

- **The Scorecard section** provides details on the scorecard methodology, the selection of factors and calibration. It also provides an outlook for future development and usage of the scorecard methodology for the Systemic Risk of Modelling.

- **The Guidance for Better Practices section** draws from the Framework and Scorecard sections, in order to offer suggestions for managing the Systemic Risk of Modelling within organisations and across the industry.

Framework

2.1 Overview

Systemic risks in modelling can arise from a number of factors. Broadly speaking, these factors can be broken down into three categories: 1) risks originating from models, 2) risks arising from organisational structures, and 3) risks originating from behaviour.

The first category deals with existence and effects arising from the limitations of models, even with perfectly rational people and organisations. This includes how data is gathered and used in models, how the model fits reality and our expectations, and more generally, how well the limitations of the model are well understood and addressed.

The second category deals with effects on the modelling process that emerge from how organisations and institutions are set up. This can range from how information flows within a company, how and when monitoring is carried out, competitive biases in the insurance or model markets, and lastly how market regulation can sometimes trade one kind of systemic risk for another. These can result in systemic risks from correlated behaviour from small sets of models being used, unwarranted confidence in models, or management not fully understanding the limitations of the modelling process.

The third category deals with the human factors that could complicate matters even with perfectly accurate models. The most common factors referred to here are human cognitive biases, shortage of training, expertise or experience, and behavioural inclinations to use models in ways that they are not designed for. Here systemic risks can emerge from people independently making the same mistakes, and being overly dependent or unchallenging of the models.

In practice there will always be a fair amount of overlap and interaction between the three categories.
For example, the availability of a limited number of acceptable models (organisational) may constrain the methods used in modelling (model) and make underwriters expect all models to behave that way (behavioural); this could also lead to resistance to using other acceptable but less popular methods that are actually more appropriate (behavioural). As a result the market could potentially have an unduly limited view of risk and a lower threshold to badly modelled risks.

Here, we discuss each factor in turn. For more full discussion and examples, see Appendix A.

2.2 Modelling Risk

"All models are wrong, but some are useful."[6]

For models to be used correctly, the limitations and assumptions of the models need to be well documented and understood. Gaining an understanding of how sensitive model Parameters and outputs are to data, comparing modelled output with expectations, and being aware of the process of handling and interpreting modelling information are examples of how a model can be used more correctly.

In this section we delineate three main factors of risk that originate from models: data risk, model risk, and process risk.

Data Risk

"Marketization and its associated models are savage masters – they push forward in a single direction. The drive to increase modellability of deals means that Property Catastrophe underwriters receive, year after year, ever more granular data, and request ever more detailed information.

The rise of vendor models has been intrinsically linked to an increase in detailed statistical data: models are both dependent on and shape the production and consumption of data."[7]

Data risk is the prospect of modelling error arising from deviations in the quality of data input compared to that expected by the model; the significance of this risk is therefore dependent on the completeness, accuracy and granularity of the data inputs compared to that used to build and validate the model, and on the sensitivity of the model to deviations of the data presented. Three kinds of data-risk in particular pose common problems:

- **Data Sources:** this represents the amount of, the interpretation, and level of access to the underlying data used to build and calibrate a model. Bad data can obviously lead to miscalibrated models, but missing data can also quietly lead to mis-estimates of risk that are hard to detect - especially when it is handled the same across companies. For instance, catastrophe models typically use large amounts of claims data to build specific damage functions, but if they treat "unknown entries" by assigning the average characteristic for a geographic area or exposure type, some classes of objects may be systematically mis-estimated.

- **Data Translation:** this represents the level of work-around and/or shoehorning required to translate the underlying characteristics into the prescribed modelling data format. There is a risk of a systematic distortion that is invisible at later stages, making people overconfident in the calibration of the model. For example, if it is not possible to encode the type of coverage because the model was originally designed for a more specific domain, disparate objects may be bundled together into an apparently homogeneous type or mislabelled as other types; this data will then potentially distort model estimates of the content of the portfolio.

- **Data Granularity:** this represents the quality of data presented to the model compared to that expected by the model. For instance, for a catastrophe model geocoding locations at low levels such as US County level would provide a very poor input or make it impossible for assessment against location-sensitive perils like flood.

6 Box, G. E. P., and Draper, N. R., (1987), Empirical Model Building and Response Surfaces, John Wiley & Sons, New York, NY. p. 424.
7 Jarzabkowski, P., Bednarek, R., & Spee, P. (2015). Making a Market for Acts of God: The Practice of Risk Trading in the Global Reinsurance Industry. Oxford University Press. p. 71

Thai Floods and CBI

Risks can arise when a location's characteristics are easily translated into the model, but certain external factors are not considered in the loss estimate. When modelling properties in a catastrophe model, contingent business interruption (CBI) is not taken into account but could be a source of substantial losses.

The 2011 Thai floods illustrate this: the heaviest flooding in 50 years affected the country severely, but in particular many industrial estates had been built on former rice fields (and hence floodplains) and were hit simultaneously. This particularly affected the global hard drive market (in 2011 Thailand accounted for 25% of the global market), and affected the Japanese economy badly. Lloyds estimated itself to be liable for £1.4 billion[i]. Here the direct damage was compounded by spreading, unmodeled CBI.

[i] http://www.theguardian.com/business/2012/feb/14/lloyds-thailand-flooding-2bn-dollars

Source https://en.wikipedia.org/wiki/2011_Thailand_floods#/media/File:Flooding_of_Rojana_Industrial_Park,_Ayutthaya,_Thailand,_October_2011.jpg

Spreadsheet errors

Spreadsheets are widespread tools in business and elsewhere, but have an astonishingly high rate of errors. According to studies[ii] close to 90% of all spreadsheets have errors – many with significant effects on business. 17% of large UK businesses have suffered financial loss due to poor spreadsheets, and far more (57%) have wasted time or made poor decisions (33%) due to spreadsheet problems.[iii] The list of horror stories is long,[iv] including budgeting errors running into billions of dollars, mistaken hedging contracts, hundreds of million dollars of client losses to investment firms, bank failures, and so on. Despite the data developers and users are often overconfident in the correctness of the spreadsheets, which contributes to the problem.[v]

[ii] Panko, R. R. (1998/2008). What we know about spreadsheet errors. Journal of End User Computing, 10, 15-21. Revised 2008. http://panko.shidler.hawaii.edu/SSR/Mypapers/whatknow.htm
[iii] F1F9, "Capitalism's dirty secret: a research report into the uses & abuses of spreadsheets" (2015) http://info.f1f9.com/capitalisms-dirty-secret
[iv] http://www.eusprig.org/horror-stories.htm , http://www.cio.com/article/2438188/enterprise-software/eight-of-the-worst-spreadsheet-blunders.html
[v] Panko, R. R. (2009). Two Experiments in Reducing Overconfidence in Spreadsheet Development. Evolutionary Concepts in End User Productivity and Performance: Applications for Organizational.

Model Risk

Additional risk can obviously result from flaws in the model itself. No amount of data or clever handling of uncertainty can compensate for an overly simplistic (or just incorrect) model. The systemic risk comes from many actors relying on models with similar mistakes, or not recognizing limitations in the same way.

The biggest components of this kind of risk include:

- **Model Diversity:** this represents the level of diversification (or concentration) in modelling error resulting from the aggregation of many modelling components. This includes the number of explicit models in use, but also the number of methodologies, internal components or alternative views that contribute to the risk estimates. Different models fitted to the same historical data and insurers using the same models will have correlated model errors, causing systemic risk if they are not countered by good model management, expertise and an individual view of risk.

- **Model Credibility:** this represents the extent of the knowledge (or uncertainty) of the real underlying phenomena captured in the model, and how much to trust it. Systemic risk can emerge when many users have inaccurate estimates, especially driven by apparent accuracy, solid data and market buy-in. For instance, when modelling very extreme but rare events (such as a nuclear meltdown, the probability of which is estimated to be of the order of one in a million per year[8]), Model Risk can become much greater than the probability of the real underlying phenomenon[9], yet the model itself and practical experience will give little indication that there is reason to doubt it.

- **Model Fitness For Purpose:** this represents the applicability of the model to the business under consideration. For instance, a vendor model can be tailored to provide a better fitness-for-purpose through an extensive validation process and associated adjustments to the model. However, some things are too complex to model, and should perhaps not be modelled in the first place since even a tailored model would be misleading. Understanding enough about the environment and intended use can help determine if modelling is fit for the purpose.

Model Diversity and Aggregation

There is an extensive literature in machine learning and statistics on 'ensemble methods'. The general idea is to train many diverse predictors on the same dataset and then aggregate the predictors to make an overall prediction. Aggregated predictions tend to outperform methods that train a single predictor. "Boosting" and "random forests" are two powerful and widely used ensemble methods.[ix] (In the Netflix $1m prize, the winners ended up being ensembles of the early leaders). An empirical survey of ensemble methods[x] suggests that most of the gains come from combining the first few predictors, but that some ensemble methods continue to get large gains up to 25 predictors.

[ix] Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10), 78-87. https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf
[x] Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research, 169-198.

8 IAEA (2004), Status of advanced light water reactor designs 2004, IAEA- TECDOC-1391, International Atomic Energy Agency, Vienna, Austria. It is worth noting that there have been three major reactor accidents in civil nuclear power (Three Mile Island, Chernobyl, and Fukushima) over 15,000 cumulative reactor-years. This suggests a rate of 2 in 10,000 per year, significantly above the modelled aim.

9 Ord, T., Hillerbrand, R., & Sandberg, A. (2010). Probing the improbable: methodological challenges for risks with low probabilities and high stakes. Journal of Risk Research, 13(2), 191-205.

Computational diversity

Combining multiple models is sometimes used in high reliability computing. In N-version programming several implementations of the same software specification are written by independent teams. These programs are then run in parallel, with output corresponding to majority decisions. Ideally the independence of implementation in this approach greatly reduces the chance of identical software faults[xv]. Unfortunately, in practice it turns out that independent programmers actually make similar errors in the same places[xvi], reducing the utility of the approach.

[xv] Chen, L., & Avizienis, A. (1978, June). N-version programming: A fault-tolerance approach to reliability of software operation. In Digest of Papers FTCS-8: Eighth Annual International Conference on Fault Tolerant Computing (pp. 3-9).
Avizienis, A. A. (1995). The Methodology of N-Version Programming, Software Fault Tolerance, Edited by Michael R. Lyu.
[xvi] Knight, J. C., & Leveson, N. G. (1986). An experimental evaluation of the assumption of independence in multiversion programming. Software Engineering, IEEE Transactions on, (1), 96-109.

Process Risk

> "Risk management is about people and processes and not about models and technology." **Trevor Levine**

Some risk from the modelling process results not from flaws with the models themselves but from their inappropriate operation within the organisation. The Modelling Process involves selecting data to build models, the actual model building, how the models are validated and used, and how the results are then used to guide action: a model can be correct but integrated in a process that produces systemic risk. As we'll see, process risk by its nature spans modelling risk proper, organisational risk, and behavioural risk.

Examples of risk factors associated with Process risk include:

- **Subjective Judgment:** this represents the level of unjustified tweaking, typically resulting from over-confidence or anchoring biases. For instance, underwriters could be anchored on irrelevant data, outdated information, or other biasing factors (see the behavioural factors section 2.4).

- **Resources, Expertise & Experience:** This is the modelling team's ability to understand the content of models and their limitations. Much of this is based on having experience with the models, both when they work as they should, when (and why) they fail, and when unexpected real world situations occur. Systemic risks can occur when teams have too limited expertise pools, for example by lacking diverse backgrounds or proper feedback, and hence tend to make similar or naive assumptions.

- **Controls & Consistency:** this represents the ability to prevent operational risks from producing unintended results from the model. While good Controls may reduce operational risks to organisations they can inculcate similar practices that make the weaknesses of the models similar across the industry. For instance, asymmetric error checking driven by the natural pressures of the market could produce systemic bias of errors (see the behavioural factors category for further detail).

Lessons from Freestyle Chess

In 'freestyle chess,' any combination of human or machine decision-making process is allowed. Currently the champions of freestyle chess are not computers, but rather human-machine combinations (called "centaurs"). In 2014's Freestyle Battle the cyborgs beat the best chess AI 53-42. Human guidance thus adds significant value. A good freestyle player knows when to listen to the computer, when to override it, and how to resolve disputes between different chess programs. Learning how to exploit the complementary abilities that the best chess programs have is therefore a distinctive cognitive skill.

2.3 Organisational Risk

"Nested rationality sheds light on the mundane practices through which models become accepted, hubris constitutes the norm, stress is celebrated, and traders become entangled in a dense relationality of miscalculation. We show that these everyday practices constitute the collective practice of the market and, hence, are the crux of systemic health – or systemic risk – within markets."[10]

To gain a fuller understanding of risk in the industry, we need to understand models not in isolation but as part of a broader system. Models are used within — and produced by — organisations embedded inside markets, subject to competitive pressures and regulatory oversight, How information flows between people in an organisation can be as important as how data is transferred between parts of a model for the eventual outcome: if transparency, correctness checks, incentives or feedback are problematic the organisation may make its members use the model in a faulty way. Competition between insurance companies can favour portfolio optimization to fit models, the use of certain models, or market cyclicity that makes company behaviour and risk more correlated. Regulation, which generally aims at reducing systemic risk, can inadvertently lead to rules that make companies behave similarly or reduce model diversity.

In this section we delineate four main risk factors arising from organisation: **Regulation, Competitive Environment, Model Markets**, and **Internal Risks.**

[10] Jarzabkowski, P., Bednarek, R., & Spee, P. (2015). Making a Market for Acts of God: The Practice of Risk Trading in the Global Reinsurance Industry. Oxford University Press. p. 192

Regulation

Firms do not necessarily have final control over which models they use and how they use them. Considering their wide reach and influence, regulators naturally have potential to create or mitigate systemic risk. Individual companies will be impacted to varying degrees depending on their particular regulator, and the impact that a regulator has will be dependent on its ability to act in a way that avoids introducing systemic risk and is acceptable to markets and other stakeholders. Just like insurance companies aim to use models to manage their risks, regulators have more or less explicit models for managing market risks - subject to exactly the same uncertainties and risks as the other risk models[11] .

Determining optimal regulatory practices vis-à-vis systemic risk is a subtle and difficult matter. For instance, if the regulator accepts a higher risk, the lower capital requirements across the market may create a systemic risk. On the other hand, enforcing high capital requirements compared to other countries may lead to firms changing domicile or becoming uncompetitive, in turn damaging the efficient operation of a sector on which the wider economy depends. Indeed, the nuance of the role regulators play in determining systemic risk is a strong motivation for developing a robust metamodel.

Risk Diversification

In addition to having a diverse set of models, firms also need to have a diversified set of risks. While minimising correlation between risk holdings is always a key consideration for an insurance firm, the need to do so becomes even more acute when considering the systemic risk of modelling. Model errors and structural uncertainty might mean that certain risks are underestimated. By diversifying risk, one can ensure not only that a single disaster wouldn't bankrupt the company, but also that a single set of model errors or poor estimates wouldn't either.

However, individual firms diversifying their assets is not always sufficient to ensure protection against systemic risk. All firms have an incentive to diversify their risk. However, diversity between firms can paradoxically diminish as all firms pursue similar diversification strategies. While individual firms are far less likely to fail with diversified holdings, the system as a whole becomes vulnerable to total collapse via these correlated risk connections. What works as a risk-reduction strategy for each individual firm could actually make the system as a whole more fragile. A series of studies by Lord Robert May and his colleagues found that this dynamic exacerbated the 2008 financial crisis. To what extent this dynamic impacts the insurance market has yet to be determined.

Uniform diversification

The inter-bank lending network is intended to reduce risk, but can under some conditions instead amplify it and produce systemic risk, such as during the 2008 financial crisis. Lord May and Nimalan Arinaminpathy modelled how liquidity shocks can spread in the network, finding how the network structure and size distribution influence its stability.

Having uniform controls can sometimes increase systemic risk. Lord May and his colleagues note that all banks have an incentive to diversify their assets. However, diversity between banks can paradoxically diminish as they all pursue similar diversification strategies. While individual banks are far less likely to fail with diversified holdings, the system as a whole becomes vulnerable to total collapse via these asset-holding connections. What works as a risk-reduction strategy for each individual bank actually makes the system as a whole more fragile.

[11] Haldane, A., & Madouros, V. (2012, August). The dog and the Frisbee. Bank of England. In Speech given at the Federal Reserve Bank of Kansas City's 36th economic policy symposium,"The Changing Policy Landscape", Jackson Hole, Wyoming (Vol. 31). http://www.bankofengland.co.uk/publications/Documents/speeches/2012/speech596.pdf

Competitive Environment

Depending on the market in which a company operates it may be subject to varying levels of competitive pressure. A suitably competitive environment can be healthy, raising standards and breeding innovation. However it can also allow negative effects to perpetuate and where a significant proportion of the market is influenced in the same way this can introduce certain systemic risks. For example, extreme competitiveness can promote willingness to use smaller relative security loadings or cut corners on capital buffers in order to attract customers, increasing individual risk but also forcing less risk-taking companies out of the market or into making similar adjustments.

The focus of this particular category are the scenarios that would lead to individual companies throughout a market to act in a way that they would not do if they were isolated from the behaviour of other companies within the market, and specifically where this both reduces individual robustness and increases the likelihood of others within the market acting in the same way.

This can be broadly considered in two dimensions:

- **Reputation:** The extent to which a company's ability to attract and retain business is linked to their reputation or external perception can dictate how susceptible they are to making decisions based on appearance, rather than with due consideration for the risk. By its nature, this effect will perpetuate through a market as each individual attempts to either follow what is perceived to be "best practice", or attempts to continually differentiate themselves from the competition. Examples include pressure to include coverage for emerging non-modelled risks, such as Cyber; or pressure to not "rock the boat" by requesting higher data quality than a competitor.

- **Irregular feedback**: In a healthy competitive P&C environment, the threat of being undercut solely on price selection is naturally corrected for through rapid feedback in the form of claims. This ensures that pure risk measures between companies cannot deviate too far from reality and forces individuals to compete on efficiency, innovation and customer service. Where irregular feedback exists, the negative effects of decisions do not occur sufficiently quickly for the action to be corrected for in this way, resulting in a consistent bias to underestimate risk in order to gain business, and a market-wide race to the bottom. Examples include extreme events that follow statistical power laws; complex risks where there are multiple interpretations of outcome; or risks with high uncertainty ranges. This effect is most extreme when the consumers are likely to be poorly informed about the nature of the risk themselves, such as rare or emerging risks - typically risks that are also badly modelled.

Getting the "right" answer

One dramatic example of how pressures to get a desired result can bias a model can be found in the JP Morgan Task Force Report, detailing how billion dollar losses were partially due to spreadsheet errors lowering the estimate of VaR in Basel II models. A mistaken formula muted volatility by a factor of two and erroneously lowered the VaR[xxi]. However, this error was very much in line with the CIO's priorities to reduce VaR in 2012, and there was great pressure to implement a new and improved VaR model which was promised to produce a lowered VaR. [xxii]

[xxi] http://files.shareholder.com/downloads/ONE/2272984969x0x628656/4cb574a0-0bf5-4728-9582-625e4519b5ab/Task_Force_Report.pdf (p. 128-129) [xxii] Lisa Pollack, A tempest in a spreadsheet, FT Alphaville, Jan 17 2013 http://ftalphaville.ft.com/2013/01/17/1342082/a-tempest-in-a-spreadsheet/.

Model Markets

In addition to being to some extent at the mercy of the regulator and the direct competition, firms have little direct and immediate control over the availability and quality of models built by third-party vendors. Models are built to follow market requirements, with the ability to write business driving the granularity and completeness of the model, rather than the underlying risk. Conversely, regions with poor data are hard to model. Measuring a company's risk exposure against the priorities of the wider market can be used to understand how aligned they are to this systemic risk.

Within a market that relies heavily on third-party models, the priorities of the vendors will be driven by market forces as with any other product. This can lead to over-emphasis on a handful of "top priority" risks, that reach the widest audience, rather than an overall adequacy of representation for the materiality of each risk to the market as a whole.

Existing model vendors do ask their customers where to improve their models, and the response is typically where (geographically or peril-wise) they have business. New vendors don't build models nobody is likely to ask for: the cost of building models means market size will determine where models are built. This could be because it is not considered material to a sufficient number of companies, or because either poor data quality or a lack of claims information increase the work required.

The result is an uneven global footprint of what is modelled, based on domestic appetite. If risks are underwritten then they tend to be modelled well, but rarer risks will be less well modelled in addition to the model uncertainty due to lack of data, and the lack of experience among underwriters in how to handle them. As a result, a significant market risk may continually be under-invested in.

Long term under-investment in a hazard model can introduce systemic risk to a market. Where no credible models exist regulators struggle to introduce appropriate monitoring; poor granularity or out-of-date models can mislead, leading to inappropriate allocation of risk across a market; and without sufficient scrutiny and feedback to continually challenge a model and improve its skill it can introduce systemic risk through behavioural effects.

Internal Risks

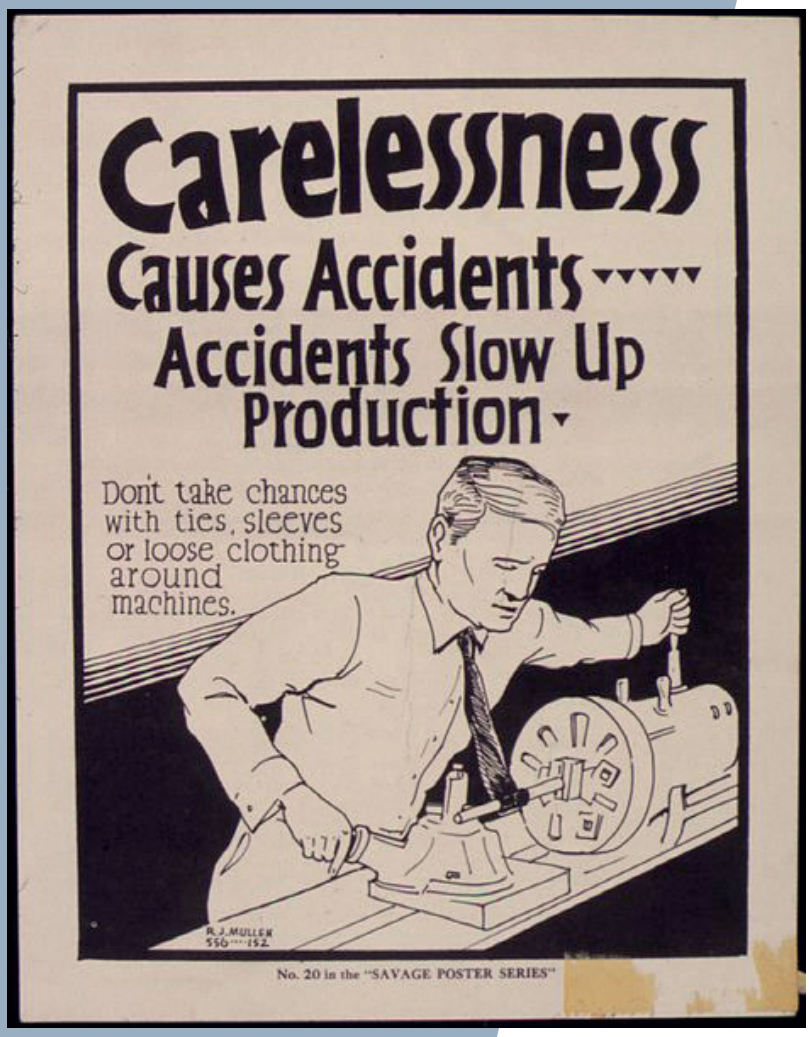
Systemic risk is introduced not only through external forces but also through the internal organizational structure of a firm. The old adage "no one ever got fired for buying IBM" has some resonance in this area. Whether seeking regulatory approval for one's model or even just sign-off from a management team, there seems little to be gained from adopting a model, approach or assumptions that you know to be distinct from your peers'. There are enough surveys by consultants that give visibility as to the range of tail risk assumptions for a model builder to be persuaded to toe the line. Ultimately, this can lead to industry-wide alignment of assumptions, generating the systemic risk of failure.

Models are now so complex that it is becoming ever more difficult for anyone at a senior level in the organisation to have a thorough understanding of its operation. The technical limitations of the model, as well as more generic model risks, can be communicated. However such limitations will not provide a binary measure that says under what circumstances model outputs should be distrusted. Rather it will suggest the situations in which its reliability or accuracy may be affected. As per the IBM adage, believing the model may be a safer move than saying that it is not to be trusted and then making a decision unsupported by the model in which your business has invested vast sums. The systemic risks of belief in model outputs was seen all too clearly during the recent financial crisis. Worse, there may exist pressures within the organisation that pushes towards accepting certain models despite safeguards.

Near Misses and Normal Accidents

When an accident happens in an organisation, it is often found that it was preceded by many earlier near misses that did not cause immediate harm. Such "normal accidents"[xxiii] are often overlooked and interpreted as signs that the system is truly resilient rather than being in danger. How a near miss is interpreted can depend strongly on whether it is experienced as a loss (it could potentially have been much worse) or not (the safeguards look adequate)[xxiv].

Handling this problem is fundamentally an organisational problem: high-pressure situations often lead to accepting risks and anomalies, but managers need to justify their assessment of near misses and people who own up to or point out mistakes should be rewarded [xxv].



Source: By Unknown or not provided
(U.S. National Archives and Records Administration)
[Public domain], via Wikimedia Commons

[xxiii] Perrow, C. (2011). Normal accidents: Living with high risk technologies. Princeton University Press.
[xxiv] Tinsley, C. H., Dillon, R. L., & Cronin, M. A. (2012). How near-miss events amplify or attenuate risky decision making. Management Science, 58(9), 1596-1613.
[xxv] https://hbr.org/2011/04/how-to-avoid-catastrophe=

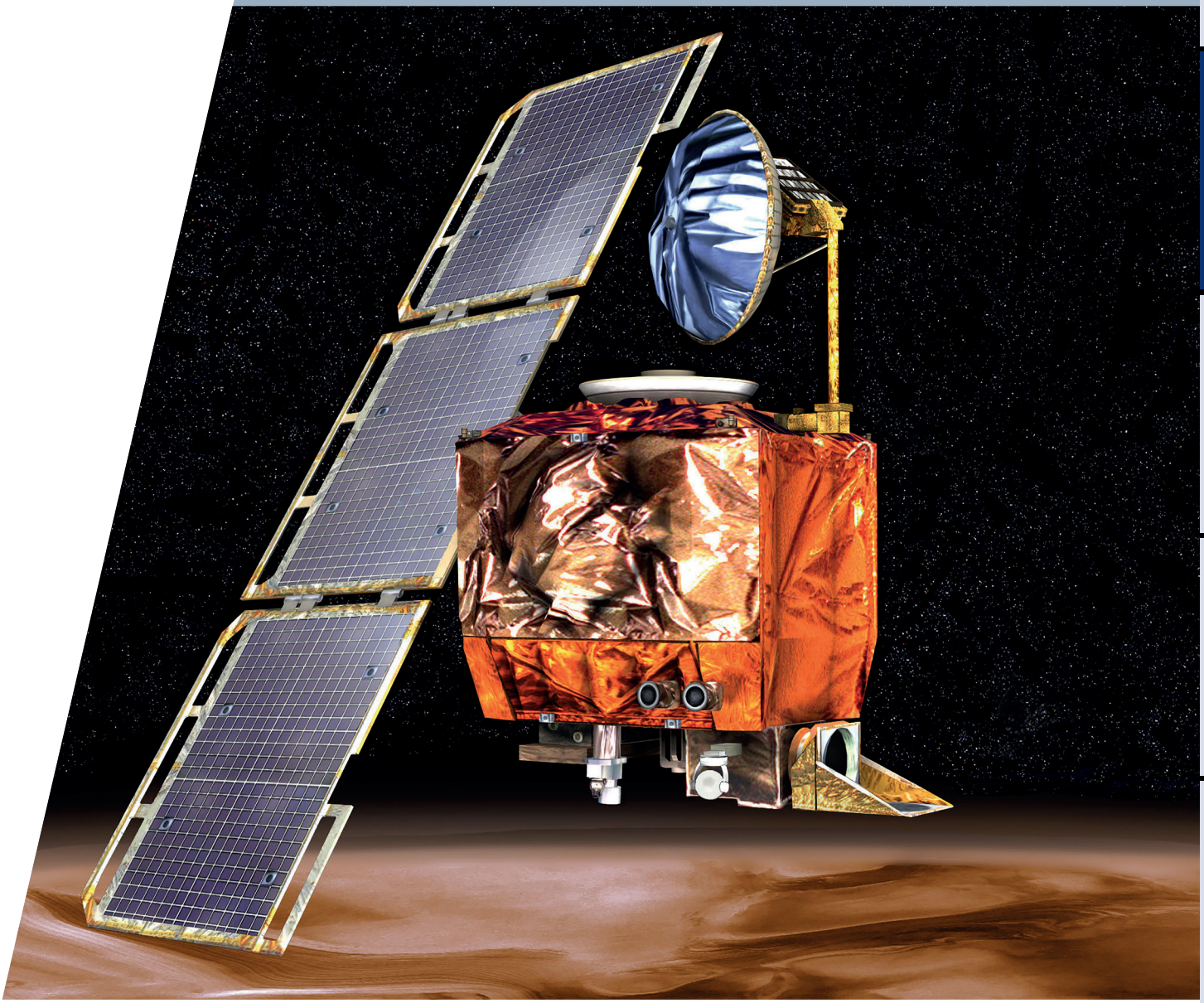
Unit conversion disaster

Mistakes in unit conversions can be costly. One of the classic examples is the, a $125 million craft that in September 1999 plunged into the Martian atmosphere and disintegrated.

The cause was found to be that one piece of navigation software used American units, but interfaced with another system expecting metric units. The discrepancies were also noted by at least two navigators whose concerns were dismissed[xxvi]. These unit conversion errors can strike critical systems, but can be avoided with observational feedback.

Models need to report enough internal information so that users can recognize that something is amiss – and have their concerns investigated.

[xxvi] Board, M. I. (1999). Mars Climate Orbiter Mishap Investigation Board Phase I Report November 10, 1999. ftp://ftp.hq.nasa.gov/pub/pao/reports/1999/MCO_report.pdf



Source: "Mars Climate Orbiter 2" by NASA/JPL/Corby Waste - http://www.vitalstatistics.info/uploads/mars%20climate%20orbiter.jpg (see also http://www.jpl.nasa.gov/pictures/solar/mcoartist.html). Licensed under Public Domain via Commons - https://commons.wikimedia.org/wiki/File:Mars_Climate_Orbiter_2.jpg#/media/File:Mars_Climate_Orbiter_2.jpg

2.4 Behavioural Risks

Because it is humans who ultimately use the models and make the most important decisions in any company, it's crucial to analyse how our own thought processes can exacerbate systemic risk. Psychologists have identified systematic biases in decision-making about probability and risk. Because we often make similar and correlated errors in judgment, there is substantial potential for the market as a whole to respond incorrectly, thereby introducing systemic risk. Generally, the insurance market will benefit by being aware of these biases, leading to a better management of risk.

In this section we analyse the Autopilot Problem and Cognitive Bias, which are two main behavioural drivers of risk. Further discussion and potential methods for risk mitigation can be found in the Experimental Section and Appendix A.

Autopilot

> "When models turn on, brains turn off."
>
> **Til Schuermann**

The autopilot problem emerges when there is a dependence on an automated system that has been introduced with the intention of replacing a human act, thus transforming the function of the human into an overseer. The autopilot problem can dictate the outcome of decisions, which in turn can either reduce or introduce system risk.

With the increased availability of high performance computing, individual companies are relying on technological advances and automated systems more than ever. Automation has played an increasingly important role in the domain of risk modelling in (re)insurance since the introduction of catastrophe models.

In order to maintain expertise, specialists must be put into situations where they receive frequent feedback about their actions (Kahneman & Klein, 2009, Shanteau, 1992). Limited or zero feedback results in individuals who are not learning or honing their skills. Older generations of pilots can sometimes ride on their skills (i.e. perform at a reduced, but still acceptable level, by using the skills and techniques acquired before the introduction of the 'autopilot' (Bainbridge, 1983)), while subsequent generations cannot develop these skills in the first place. Expertise and experience can dictate the outcome of decisions, which in turn can either reduce or introduce systemic risk to a company.

Source: "PKIERZKOWSKI 070328 FGZCP CDG" by Pawel Kierzkowski - Own work. Licensed under CC BY-SA 3.0 via Commons -https://commons.wikimedia.org/wiki/File:PKIERZKOWSKI_070328_FGZCP_CDG.jpg#/media/File:PKIERZKOWSKI_070328_FGZCP_CDG.jpg

Air France Flight 447

In 2009, the autopilot of Air France Flight 447 disengaged suddenly while the plane was on the way from Rio to Paris. This was triggered by a technical failure, but the pilots couldn't know this, as they had not been actively flying the plane to that point. They were trying to simultaneously figure out what was going wrong and correct it without the necessary comparison data. The pilots never knew what the problem was before the plane hit the waters of the Atlantic killing everyone on board. In order to maintain expertise, one must be put in situations with frequent feedback. Lacking this feedback, expertise and skills degrade.[xxviii] [xxix]

[xxviii] Kahneman, D., & Kelin, G. (2009). Conditions for intuitive expertise: a failure to disagree. American Psychologist, 64(6), 515-526.
[xxix] Shanteau, J. (1992). Competence in experts: The role of task characteristics. Organizational behaviour and human decision processes, 52(3), 381- 410

The autopilot bias in this context can be mediated by a number of factors, five of which we explore further in Appendix A:

1. **Skills and knowledge** of those using the model. Can underwriters estimate risk without a model beforehand?
2. **Integration and reliance** of model usage. To what extent are those using the model reliant on the model, and to what extent is the model integrated into a more well-rounded risk assessment framework?
3. **Calibration and overconfidence** of those using the model. How well are underwriters calibrated to areas the model doesn't cover, and are they overconfident?
4. **Modelling narrative** of decisions and reporting. Do qualitative interpretations of model output accompany the risk estimates?
5. **Company culture** surrounding the use of models. How good is communication between underwriters and modellers? What role do models play in the company?

Cognitive bias

Cognitive biases are systematic deviations in human thinking from ideal reasoning patterns. Many biases are the result of cognitive limitations (bounded rationality) and can be adaptive: in many environments, they lead to quick and approximately correct answers. Unfortunately, in many modern day situations that our ancestors didn't face—such as those that require careful attention to precise probability and risk—our judgment can be led severely astray.

> "A general principle underlying the heuristics and biases is that humans use methods of thought which quickly return good approximate answers in many cases; but also give rise to systematic errors called biases."
>
> **Eliezer Yudkowsky**[12]

Although much of the cognitive work of insurance has been outsourced to models better equipped to handle large data sets than humans are, underwriters must rationally assess modelled output, risk, and unmodelled exogenous factors to make decisions. These sorts of decisions are, unfortunately, highly susceptible to errors due to cognitive bias.

Cognitive biases play a large role in how individuals and companies ultimately make decisions. Indeed, behavioural economics and behavioural finance are entire academic fields largely devoted to studying their effects. It's then quite important to understand how these psychological factors work, how they affect risk, and how we can mitigate them.

A complete discussion of cognitive bias is beyond the scope of this study, but we briefly describe some primary and representative risks with further elaboration in the next section and in Appendix A.

[12] Yudkowsky, Eliezer. "Cognitive biases potentially affecting judgment of global risks." Global catastrophic risks 1 (2008): 86.

1. **Anchoring**

Humans tend to fix their estimates of a given quantity on a single piece of information, even when that information is known to be unrelated to the value of that quantity. For instance, in one experiment, subjects were asked to write the last two digits of their United States Social Security Number and then asked whether they would pay this number of dollars for an item of unknown value. When asked to bid on these items later, subjects with higher numbers submitted bids between 60% and 120% more than those with lower numbers. The social security numbers thus served as anchors which biased the ultimate value estimate.[13] Do risk assessors excessively anchor on irrelevant model output?

2. Availability Heuristics

The availability heuristic is our tendency to estimate the likelihood of an event based on how mentally available similar examples are.[14] Imagine, for example, that you're asked to determine whether a random English word is more likely (a) to begin with the letter K or (b) to have K as the third letter. According to a study by Kahneman and Tversky, subjects tend (incorrectly) to answer that (a) is more likely.[15] The reason is that it is much easier to think of words that begin with K than to think of words that have K as the third letter. That is, words starting with K are more mentally available to us, and we use their availability as a guide to their frequency. Because the emotional intensity of an event affects its ease of recall, personal experience often plays an undue role in judgment.[16] We hypothesize that underwriters who experienced losses from Katrina are more risk averse with regard to wind policies than are younger underwriters, despite the fact that both groups have access to roughly the same data.

3. Bystander Effects

The *bystander effect* refers to the tendency people have not to intervene when others are present. For instance, in emergency situations, it often takes longer for a victim to receive help when surrounded by a large group of people as opposed to asking one bystander for help.[17] This effect carries over to non-emergency situations as well. In particular, if a worker notices a problem with, say, how the company uses a model, he may be less likely to voice his concerns if a number of others could also speak up. Research has shown that a number of factors contribute to this behaviour, but two of the most relevant for the purposes of insurance include (1) the problem of interpretation, and (2) diffusion of responsibility. The first refers to the bystander's possible doubt that he has actually identified a real problem. Since others are in the same position as he is, he might think that if there really were a problem, somebody else would have already said something. Furthermore, if he actually voices his concerns and turns out to be mistaken, he might look foolish to the group. The second refers to the fact that, when there are multiple bystanders, no individual necessarily feels responsible. Everybody else, in his eyes, is equally responsible, so he becomes less likely to intervene.

4. Biased Error Search

When imperfect models and data sets exist to estimate risk, underwriters and modellers make decisions about when to search models and data sets for errors, what types of errors to look for, and when to stop looking in a way that tends to vindicate pre-existing views about risk. This pattern of biased error search is partially driven by confirmation bias and automation bias. Under time constraints, biased error search leads to finding more expected errors, and fewer unexpected errors but and more errors in total. Quantitative information about these trade-offs is unknown, but the trade-offs could be substantial.

[13] Ariely, D., G. Loewenstein, and D. Prelec (2003). "'Coherent arbitrariness': Stable demand curves without stable preferences." Quarterly Journal of Economics 118(1): 73--105.
[14] Kahneman, Daniel and Amos Tversky (1972). "Subjective probability: A judgment of representativeness." Cognitive Psychology 3: 430-54.
[15] Kliger, Doron, and Andrey Kudryavtsev. "The availability heuristic and investors' reaction to company-specific events." The Journal of Behavioral Finance 11.1 (2010): 50-65.
[16] Stalans, Loretta J. "Citizens' crime stereotypes, biased recall, and punishment preferences in abstract cases: The educative role of interpersonal sources."Law and Human behavior 17.4 (1993): 451.
[17] Darley, John M., and Bibb Latane. "Bystander intervention in emergencies: diffusion of responsibility." Journal of personality and social psychology 8.4p1 (1968): 377.

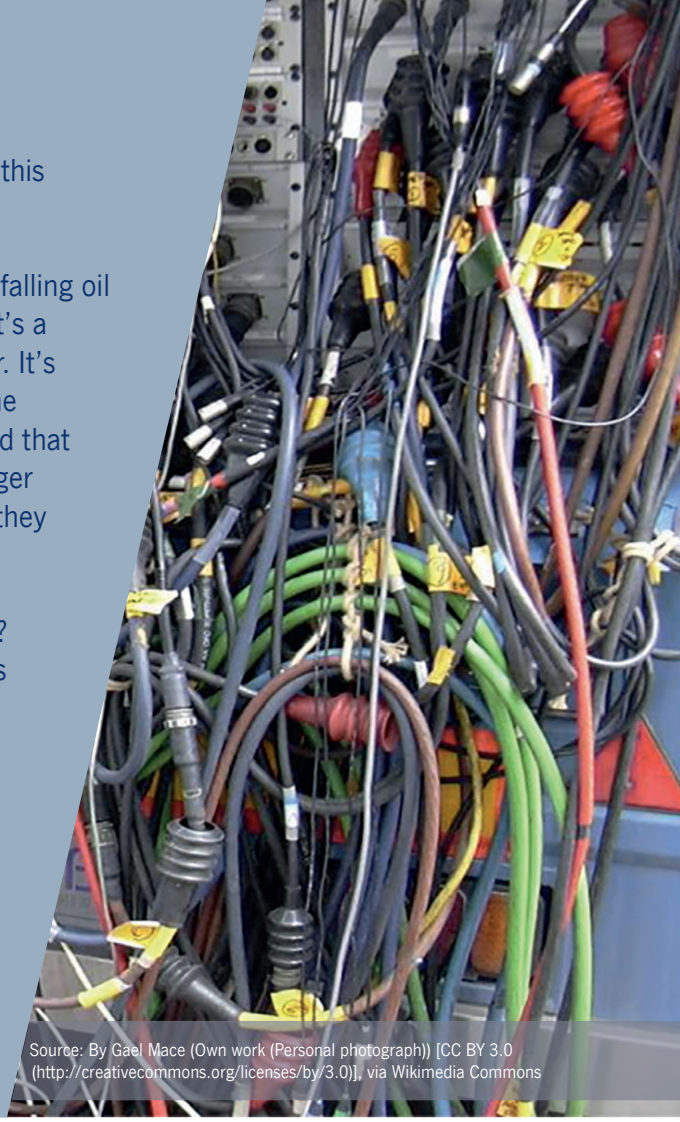
Biased Error Search

Biased error search is a familiar problem in science, as illustrated by this example from Richard Feynman:

"Millikan measured the charge on an electron by an experiment with falling oil drops, and got an answer which we now know not to be quite right. It's a little bit off, because he had the incorrect value for the viscosity of air. It's interesting to look at the history of measurements of the charge of the electron, after Millikan. If you plot them as a function of time, you find that one is a little bigger than Millikan's, and the next one's a little bit bigger than that, and the next one's a little bit bigger than that, until finally they settle down to a number which is higher.

Why didn't they discover that the new number was higher right away? It's a thing that scientists are ashamed of—this history—because it's apparent that people did things like this: When they got a number that was too high above Millikan's, they thought something must be wrong—and they would look for and find a reason why something might be wrong. When they got a number closer to Millikan's value they didn't look so hard. And so they eliminated the numbers that were too far off, and did other things like that."
(Feynman 1974, p.12) [xxxi]

[xxxi] Feynman, R. P. (1998). Caltech's 1974 commencement address. Reprinted in Feynman, R. P. (1998). 6. Cargo Cult Science. The Art and Science of Analog Circuit Design, 5



Source: By Gael Mace (Own work (Personal photograph)) [CC BY 3.0 (http://creativecommons.org/licenses/by/3.0)], via Wikimedia Commons

2.5 Risk Factors Summary

Systemic risk can come from a numerous sources. Here, we have decomposed the factors of systemic risk into three primary categories: risks arising from models, organisational structures, and behavioural factors. Within each of these categories, a wide range of sources can contribute to systemic risk, ranging from poor data quality to motivated cognition. The next stage of the research sought to determine to what extent each of these factors contributed to overall systemic risk, and take a first step to designing a countermeasure in the form of a systemic risk scorecard. The goal is to enable decision makers to monitor and assess systemic risk, and modify their behaviour to mitigate their contributions to such risk.

Scorecard

3.1 Introduction and purpose of a scorecard

The working party has developed a scorecard, which is useful for risk managers to form a preliminary snapshot of monitoring potential systemic risk. The aim is less towards an exact risk measure and more towards an indicator of whether modelling practices are aligned with reducing systemic risk. The scorecard combines different individual contributing factors that are explained in the chapters earlier into a summarised single score. The weight of each individual factor is estimated by a calibration process that consists of: (1) an analysis of understanding underwriters' pricing behaviour in the real world; (2) computational simulations based on the Metamodel; and (3) a structured expert opinion elicitation based on Delphi method. (Detailed descriptions are included in the Appendices B and C).

> "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."
>
> **John Tukey**[18]

In our opinion, the scorecard is intended as a rough way of monitoring systemic risks, not so much for decision-making, marketing, or prediction, but mainly for nudging users' behavioural changes: Going through the scoring exercise will ideally force the participants to consider the modelling process from different perspectives, becoming aware of the peculiarities of the process and where weak points in it will always be more useful than any statistical score. In fact, systemic risk is literally everybody's problem and being aware of the issue and ready to reduce the risks where possible is both the moral and practical thing to do. As discussed before, systemic risks are risks that emerge from the way parts of a system are assembled and used rather than the individual parts themselves. This can happen on multiple levels: the parts of models interact to produce a problematic model output, the (mis)use of models inside insurance companies can produce bad business decisions, and correlated model usage can make entire markets more vulnerable than they look. The scorecard helps users to take this into account by looking at factors belonging to the different levels and combining them into the final score: even if one has a perfect model usage in one's own company, one can be exposed to systemic risks from the surrounding market.

Nevertheless, the scorecard is "local". What the scorecard attempts to capture is a sense of the contribution to overall market systemic risk due to a company's modelling practices. While one can imagine evaluations of entire markets carried out by regulators, where we think the scorecard can actually be useful is for self-evaluation of companies using their own local information. This is plausible because there appears to be a fair correlation between modelling practices likely to cause risk to individual companies and to the entire market, and since proper systemic risk reduction work begins at home.

3.2 Design and elements of scoring system

Selection of factors

The first aim for the working party was to select factors that affect systemic modelling risk. It temporarily divided into three workstreams in order to more deeply investigate risk factors linked to model, behavioural and organisational systemic risks. After separate discussions a set of factors was selected with reasonably low overlap. These were further refined in workstream sessions, producing the current list. In parallel, discussions about observable signs of the risk factors and appropriate weighting systems were made.

18 Tukey, J. W. (1962). The future of data analysis. The Annals of Mathematical Statistics, 1-67.

Not all factors are equal. Some were found to likely have smaller effects than others, or were very hard to estimate. In addition, a scoring system attempting to cover all factors discussed in this whitepaper would be prohibitively complex and cumbersome. We hence selected a smaller subset aiming at capturing those that tended to come out on top in the working party discussions, literature surveys, and simulations. Some of the factors used in the scorecard are composites of, or correlated with, several of the factors discussed before, combining more information into a simple score. Another important consideration is controllability: some risk factors may be set by regulation or the nature of business and are not amenable to easy improvement. They may nevertheless be worth noting even if they affect every market participant in the same way.

Scorecard Design

The scorecard works by summing weighted ratings of all selected factors, which are themselves linearly weighted averages based on observable measures, to produce an overall systemic risk score.

Obviously nonlinear models could be made, for example taking question answers as inputs to a neural network or a nonlinear regression (as is sometimes done in credit scoring). However, the simple linear model is robust, allows investigation of what answers caused particular results, and is widely used where nonlinear models would be hard to motivate. In fact, there might be a principled case for linear models given that human judgement is often surprisingly well modelled (or even outperformed) by linear models[19].

This is particularly true when inputs have a monotone effect on the output, errors of measurement or estimation are likely, and the model is not overly sensitive to weighting[20].

Setting the weights of factors can be carried out in different ways. Had extensive statistics been available for systemic risk failures due to modelling been available, it would have been possible to use a regression fit – but this kind of data is scarce. At the opposite end of complexity lies simple tallying: in many situations even non-optimal weight setting using equal weights ("improper linear models") produces good results[21] , especially when experts choose the factors that are informative and matter in practice.

The working party combined several methods for getting estimates of factor importance. A structured expert opinion elicitation based on Delphi was used to both estimate the degree of initial group consensus, and to develop an informed view (see Appendix C. Other input came from the Oxford Metamodel of the role of modelling in the insurance market that had been developed at the Future of Humanity Institute (FHI). In particular, the metamodel allows for a way of estimating how much systemic risk changed due to different factors, which was used to estimate their relative weight in the scorecard (see Appendix B). A final input was experimental data from a FHI pilot study of cognitive bias in underwriters. Combining these sources produced the final estimated weighting of the factors in Figure 3.

[19] Goldberg, Lewis R. 1968. Simple models or simple processes? Some research on clinical judgments. American Psychologist, 23:7, 483-496.
[20] Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. Psychological bulletin, 81(2), 95.
[21] Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. American psychologist, 34(7), 571.

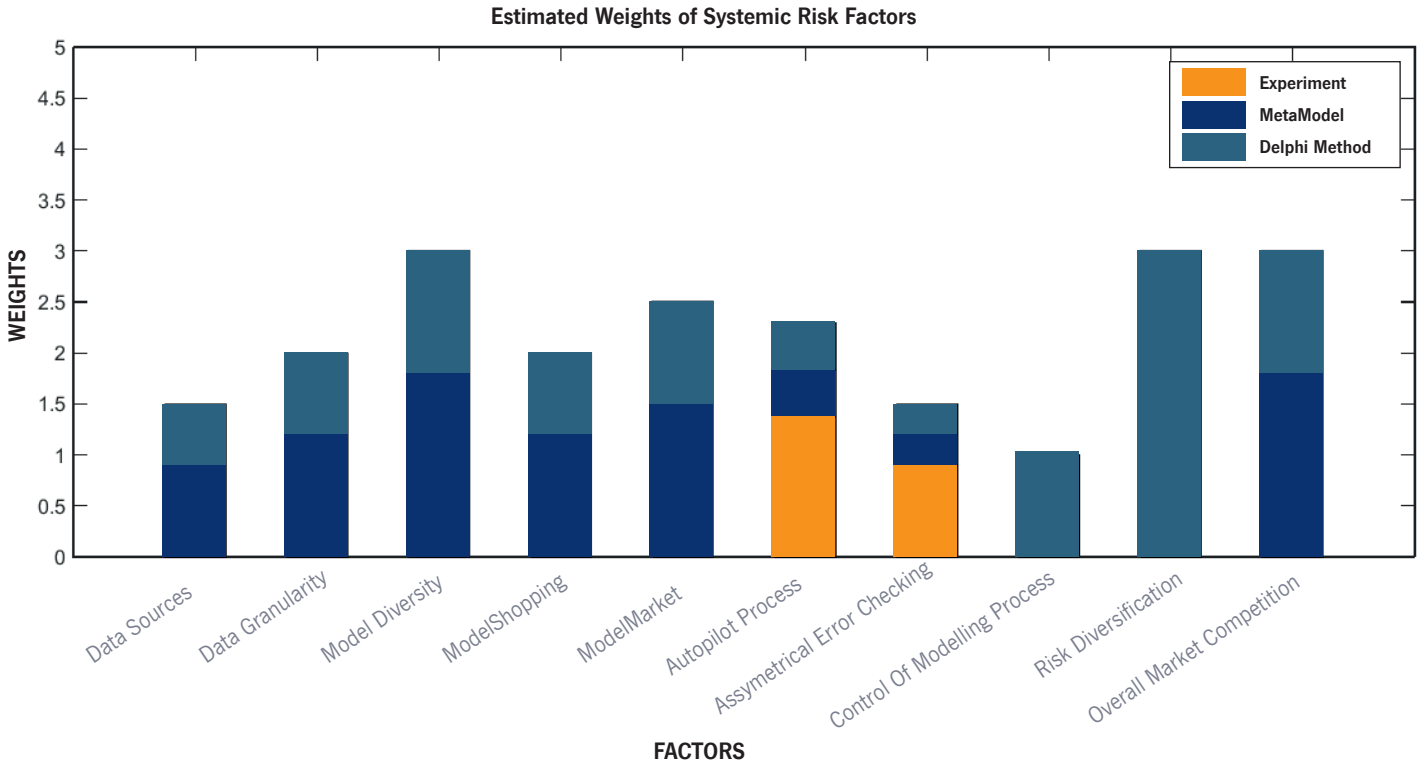


Figure 3: relative weight/importance of the factors on the scorecard.

Based on these calibration methods, a summary of (relative) weights for the selected top 10 factors is in the Table [1]. For each question multiply the individual answer by the question weight and sum it all together, dividing by 5. The total max score is 100, corresponding to maximal systemic risk.

The importance is not so much the absolute numbers as that they give information about where efforts may be needed.

Reference	Key Factors and Metrics	Weight	Observerable Measures	Measurements (Score 0-5)	Your Score
1	Data Sources	7	What proportion of your data comes from first-hand, reliable sources?	0=100% 1=80% 2=60% 3=40% 4=20% 5=0%	
2	Data Granularity	9	What proportion of your data deviates from the optimal granularity, requirements expected by the model?	0 = 0% 1=20% 2=40% 3=60% 4=80% 5=100%	
3	Model Diversity	14	What is the spread of the contribution of different parts of the model to the modelled results? This is estimated by "Model entropy". A low entropy model would have all of the result depend on a single part, a maximum entropy model would weight them all equally. If xi is the contribution of part i (where the sum of the xi is 1), the normalized entropy is: $H = \left(\frac{-1}{\log(N)}\right) \sum_{i=1}^{N} x_i \log(x_i)$ This entropy is between 0 (total concentration from one part) and 1 (even distribution)	0 = Entropy of 1 (even contribution from all parts) 1 = Entropy of 0.9 2=Entropy of 0.8 3=Entropy of 0.7 4= Entropy of 0.5 and 5 = 0 Entropy (all contributions from one part	
4	Model Shopping	9	What proportion of your model selection decisions are based on appropriateness, fitness for purpose and scientific credibility of the model? As opposed to other considerations such as price, regulatory approval, market acceptance, global licenses and relationships.	0=100% 1=80% 2=60% 3=40% 4=20% 5=0%	
5	Model Market	11	What proportion of your model or methodologies are subject to a restricted pool of suppliers / methodologies (defined as less than 3 suppliers/methodologies)?	0=0% 1=20% 2=40% 3=60% 4=80% 5=100%	
6	Autopilot Process	10	What proportion of the management and key metric information from the model is produced at a frequency which does not allow for review, narrative and challenge (e.g. weekly or more frequent)?	0=0% 1=20% 2=40% 3=60% 4=80% 5=100%	

Reference	Key Factors and Metrics	Weight	Observerable Measures	Measurements (Score 0-5)	Your Score
7	Asymmetric Error Checking	7	Given a decision, what is the probability that it will be checked carefully even if it is apparently "normal"?	0 = at least 10% 1 = 5% 2 = 1% 3 = 0.5% 4 = 0.1% 5 = less than 0.1%	
8	Control of Model Process	5	What proportion of your modelling process is subject to a well governed, controlled, effective and documented control framework?	0 = 100% 1 = 80% 2 = 60% 3 = 40% 4 = 20% 5 = 0%	
9	Risk Diversification	14	What is the spread of the contribution of different risks to the modelled results? Calculated by "Risk Portfolio entropy". A low entropy portfolio would have all of the result depend on a single risk, a maximum entropy portfolio would weight them all equally. If xi is the contribution of risk i (where the sum of the xi is 1), the normalized entropy is $H = \left(\frac{-1}{\log(N)}\right) \sum_{i=1}^{N} x_i \log(x_i)$ This entropy is between 0 (total concentration from one risk) and 1 (even distribution)	0 = Entropy of 1 (even contribution from all parts) 1 = Entropy of 0.9 2=Entropy of 0.8 3=Entropy of 0.7 4= Entropy of 0.5 and 5 = 0 Entropy (all contributions from one part)	
10	Overall Market Competition	14	What proportion of your business is from highly competitive markets where the market price can deviate by more than 30% below the technical price within the prevailing underwriting cycle?	0 = 0% 1 = 20% 2 = 40% 3 = 60% 4 = 80% 5 = 100%	

Table 2: suggested weights and possible observable measures for defining factor values

Scoring system calculation (An administrative example of scoring system administration)
This is a fictional example of how the scoring system may be used by a company. The example is specific to Catastrophe Modelling, however the scorecard methodology may be applied to all types of insurance modelling.

Data sources: For the first year, the company estimates that about 60% of its data comes from first-hand sources, giving measurement of 2. Multiplying by 7 for the factor weight, the contribution to the final score is 14. In the second year, because of a shift away from reinsurance data, about 80% of data is close to first-hand, and the measurement is now 1.

Data granularity: about 40% of business is written with only county level data, while it is known that the peril in question (for example flooding) is best modelled at a finer granularity. This produces a measurement of 2.

Model diversity: the company writes business using three main models, where one of the models has five times more influence (in terms of importance for decisions and amount of business written; for example, it may always be used, while the other two models are only consulted for special cases). This produces x1=0.14, x2=0.14, x3=0.71 and entropy of 0.72, giving a measurement of 3.

Model shopping: The Company writes about 40% of its business using a model imposed largely by limited model choice, giving a measure of 3. By year 2 this has improved because of availability of new, apparently scientifically credible models, and the measure becomes 2.

Model market: in year 0, there is essentially only one available model supplier for the relevant market, producing a measurement of 5. In year 1, new models have arrived applicable to 40% of the business, reducing it to 3.

Autopilot: About 60% of the information from the model is produced at a high frequency that makes it hard to criticise, producing a measure of 3.

Asymmetric error checking: The internal quality processes of the company spot checks about one in a thousand cases, producing a measure of 4.

Control: Due to the changes in model usage and what business is written, the documentation of the modelling process slips somewhat, increasing the measure from 2 (60% usage is well documented) to 3 (40%).

Risk diversification: Because of market changes, the risks become less spread between different types. Originally the total risk portfolio was 12%, 13%, 13% and 62% (entropy 0.77), but in year two a focus on the last two forms of risk changes it to 3%, 7%, 15%, and 75% (entropy 0.52) increasing the measurement to 4.

Market competition: Much of the company business is in a market where prices often reflect preferences for customer retention, producing a measure of 3.

These measurements are multiplied by factor weights, summed, and normalized to a scale 0-100.

Selected Factors	Weight	Measurement (Year 0)	Weighted Measurement	Measurement (Year 1)	Weighted Measurement	Change of Systemic Risk
Data sources	7	2	14	1	7	-7
Data granularity	9	2	18	2	18	0
Model diversity	14	3	42	3	42	0
Model shopping	9	3	27	2	18	-9
Model market	11	5	55	3	33	-22
Autopilot process	10	3	30	3	30	0
Asymmetric error checking	7	4	28	4	28	0
Control of modelling process	5	2	10	3	15	5
Risk diversification	14	3	42	4	56	14
Overall market competition	14	3	42	3	42	0
Total	**100**		**308/5=61.6**		**289/5=57.8**	**-19/5=-3.8**

In this case we can see that the company is scoring in the high middle of the scale between perfect (0 score) and worst possible (100): there is room for improvement. Compared to year zero the company has improved its systemic risk, mostly by the change in model market and a better model shopping avoidance strategy. However, this is offset by worsening risk diversification: the shift in business may have reduced the problem of data quality at the price of focusing too much on particular risk areas.

3.3 Use of scorecard and outlook

Clearly, a scoring system is an approximation to reality: it maps a complex domain into a simple estimate. It will not function well if the data or theory it is based on is not representative or correct enough. While we have good confidence that this scoring system points in the right direction, is simple enough not to suffer elaborate overfitting, and would apply to a wide range of modelling, there will always be cases where it cannot apply. A scoring exercise should always include considerations of "Does this question make sense in our context?" - understanding how one's business differs from the normal is important for having a proper view of one's own risk. By context, the user should think of the modelling environment in the wider context: what are the modelled outputs used for? Are systemic risk drivers inherently challenged as a result of the organisational structure? Are there incentives to introduce bias in the model use?

This also matters for comparability. The interpretation of the factors will by necessity be tied to the nature of the company and its business, and may hence change over time. Just because two companies score the same does not mean that they have the same type of risk, or that they would agree on whether the overall systemic risk is at an acceptable level. A regular user will soon consider how to make the scorecard tailored to the modelling environment within the company, adding or modifying the factors to reflect new understanding of the skill of the models, just as for normal risk it is important to own one's own view of systemic risk.

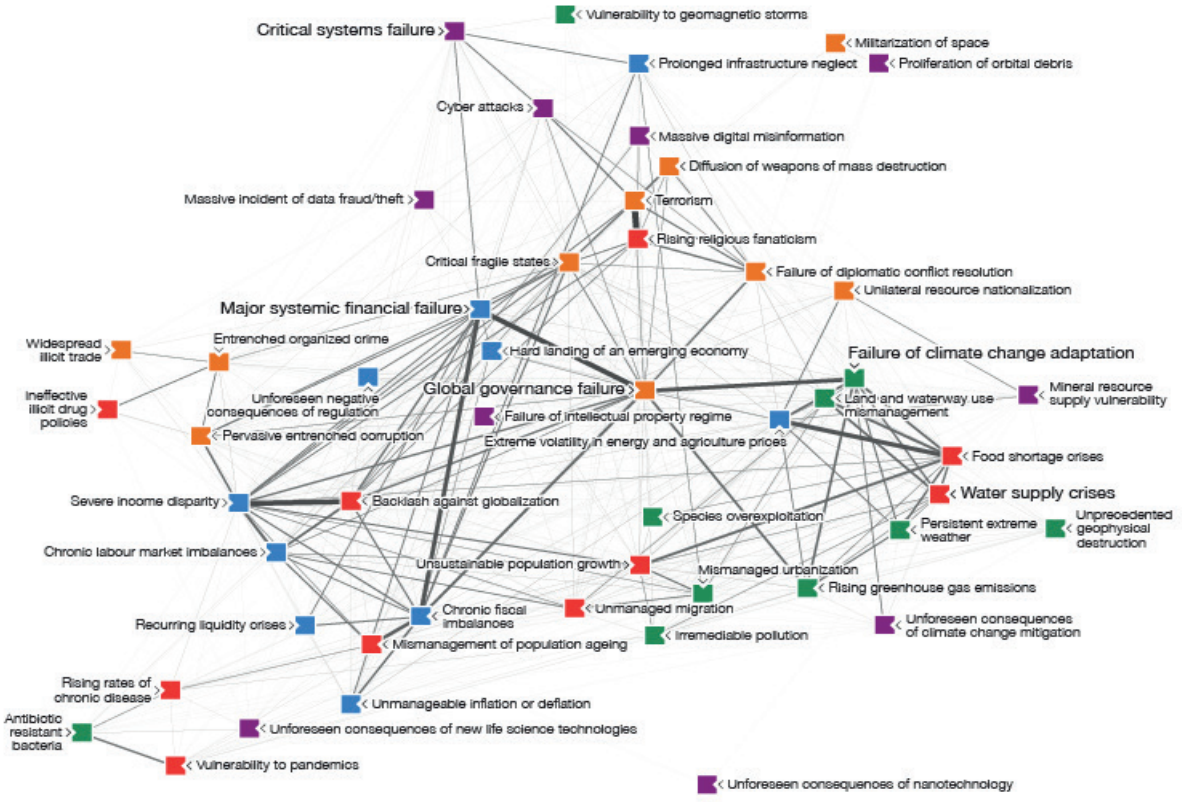
Last but not least, an important issue is the so-called "Goodhart's law", most popularly described as: "When a measure becomes a target, it ceases to be a good measure." People anticipate the response, and begin to game the measure rather than try to achieve the end for which the measure was invented. A relevant restatement is that risk models commonly break down when used for regulatory purposes.[22] Using the scoring system for regulation would definitely be unwise even if it had perfect statistical and theoretical rigor, since it would no doubt be gameable. This is why we advise against using it for comparing organisations or doing decision-making. It is better at helping discover what can be improved than giving the proper incentives for improvement. The scorecard in its present state is just the first step towards proper systemic risk measurement and sustainable modelling. We hope it will serve as inspiration for better tools in the future. It can and should be improved in many ways: through feedback and critique from users, more detailed experimental, simulation and expert input, by being compared to actual market data over time, and deeper investigations into the nature of SRoM.

There is an important difference between being aware and observing a risk, and being able to mitigate it efficiently. The scorecard helps with the first half, but better mitigation strategies are needed. It also does not cover risk triggers: mapping out what would trigger cascades of systemic mistakes would further help guide mitigation.

Expanding the investigation to the insurance linked security market and other uses of risk models may also prove useful: such instruments have begun to connect previously uncorrelated markets (insurance and capital) in ways that may pose systemic risks. Another potentially fruitful area is understanding how model-makers, individual firms and (quasi-)regulators can coordinate to reduce joint systemic risk, one where a shared model or score for systemic risk would be helpful just as modelling has been as a shared language in insurance itself.

[22] Daníelsson, Jón (July 2002). "The Emperor Has No Clothes: Limits to Risk Modelling". Journal of Banking & Finance 26 (7): 1273–96.

Summary guidelines for better practice



Source: World Economic Forum

4.1 Systemic risk in the larger world

Systemic risk is increasing because of the growing globalisation, interconnectedness, and speed of our world[23]– trends that are generally beneficial, but introduce new risks. Modelling is also going to become increasingly common: the natural response to a complex risky world is to try to manage it, and our modelling capabilities will increase radically over the coming years thanks to better computers, automatically collected massive datasets, and new data science methods.

Done naively this will create a perfect storm of modelling systemic risk, where shared models used by overconfident actors lead global markets into more risk[24].

However, learning the lessons from past mistakes and planning ahead about systemic risk can help us avoid this. Systemic risk is by its nature often difficult to notice until a catastrophe strikes. Because it depends on how various parts of the system interact even if it is noticed it can be very hard to mitigate.
On the positive side, different parts can also support each other to reduce the risk: we can design our modelling practices, organisational cultures and markets to mitigate systemic risk.

[23] Goldin, I., & Mariathasan, M. (2014). The butterfly defect: How globalization creates systemic risks, and what to do about it. Princeton University Press.
[24] Haldane, A., & Madouros, V. (2012, August). The dog and the Frisbee. Bank of England. In Speech given at the Federal Reserve Bank of Kansas City's 36th economic policy symposium, "The Changing Policy Landscape", Jackson Hole, Wyoming (Vol. 31).

4.2 Regulation, policy, practice

Monitoring of Systemic Risk of Modelling (SRoM) at industry level

Systemic risk is often a "Tragedy of the commons" problem, in that Individual rational actors can act in ways that produce a shared problem. This typically requires coordination to solve. In the case of SRoM the first step should be monitoring the risk on the industry level, since this can both help estimate how much mitigating effort is needed individually and jointly. This will require the support from impartial trusted third parties (such as Lloyd's) and/or regulators to handle issues of information sharing.

Stress testing for SRoM at industry level (e.g. major flaw in major model)

One useful experiment would be stress testing for SRoM at the industry level, for example by running an exercise considering the effects of a major flaw in major model. This can also help quantify the actual benefits of model diversity and the overall systemic risk due to present practice.

Regulatory disclosures on SRoM

Regulatory disclosures on SRoM (along the lines of the risk factors highlighted in section 2) may be helpful. At present regulators mainly look for systemic risks due to the more traditional financial sources of contagion and diversification. Finding a useful form of SRoM disclosure is a challenge: merely stating scores of the SRoM scorecard is not enough. Systemic risk disclosures are fundamentally qualitative, but it is easy to turn reporting into a box-checking exercise rather than actually providing useful information[25]. The aim at this time should likely be to begin the process of understanding what would be useful and how it could be done.

4.3 Making more resilient organisations and markets

Model Independent Scenario Analyses

Systemic risk increases if all inputs to decision-making are model-dependent or filtered through the same cognitive biases. To increase resilience the organisation can develop model-independent approaches to support their decision-making, for example relying more on raw information, Stress Scenarios or Realistic Disaster Scenarios (provided they do not rely on modelling output), or accumulation analyses like "spider bombs".

These approaches tend to be less refined, but they can complement the modelling approach and potentially flag any over-reliance on the model. For instance, the modelling output may indicate very little risk because it assumes very low probability of occurrence; but the accumulation analyses show that there are huge exposures and the losses would be huge if the model proves to be wrong. The lack of refinement is also a safeguard against overfitting and biased model selection.

[25] One challenge is to avoid Campbell's law: "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor".

Training

Pilots train for manual landings and equipment failure; it may be useful to consider training for handling model failure, both total and partial. This is locally useful inside the organisation to maintain skills and critical thinking about models, and across a market to reduce overall systemic risk.

One particular method that has been suggested to avoid getting focused on specific (possibly spurious or model dependent) probabilities or ignore badly behaved tail probabilities is to consider what scenarios can actually be handled by the company. For example, doing reverse stress-testing to see what the least extreme scenario that could destroy the business model of a company, has the potential to find specific scenarios that are both actionable and give a sense of what truly is at risk[26].

Learning from close calls

Extreme tail risks by definition rarely occur, making models of them unreliable. This is doubly true for SRoM, since extreme model failures may be unprecedented (models may have not been used long in the market, and the market itself is changing). However, close calls when recognised and interpreted as warnings (rather than reassurances that the safeties are adequate) do give some information about systemic risks. Building organisations and markets that can pick up this information can strengthen resilience. It requires procedures and a culture that recognizes the utility of disclosure when somebody notices a problem, and on-going sceptical evaluation of what is going on even (or perhaps especially) when things seem to work well.

Maintaining model diversity

Extensive practical testing in the fields of machine learning and statistics suggest that predictions can be improved by fitting multiple models and combining their predictions. Such 'ensembles' will be more useful to the degree that individual models miss important aspects of the domain being modelled. Similar results from the Oxford Metamodel (as well as the intuition of experts) suggest that increasing the diversity of models between firms reduces systemic risk for the industry. Although individual firms can acquire comfort from using models that are generally accepted in the market conforming to a common view, firms are likely to feel competitive pressure to use certain models similar to their competitors. This is a particular instance of the general conflict between good practice for individual firms and good practice for the health of the entire industry. An additional source of pressure to conform comes from regulatory requirements that are easier to satisfy with industry-standard and accepted models. Cooperation between regulators and the industry on modelling issues could result in wider appreciation of model systemic risk. Greater tolerance of justified divergences in estimates could help maintain a diversity of views and hence reduce systemic risk, which should be as important a regulatory goal as the solvency of individual participants.

A diversity of models, while beneficial, is not a panacea. Choosing a model for a particular domain is a challenging task and models have to be validated and carefully adjusted before use. An additional source of risk is the lack of a shared understanding of key properties of the model between the modellers, the underwriters and the upper-level management.

The importance of diversity extends beyond the models themselves. When it comes to systemic risk, models can only be so helpful. Detailed modelling of tail risk may not be possible due to limited data. Models that are diverse along many other dimensions may end up making the same incorrect prediction about a tail event. For this reason, it's important that the limitations of particular models are understood, and that underwriters are able to consider diverse scenarios that might be poorly captured by their models.

[26] Mary Pat Campbell, M.P. (2012). Minimally Destructive Scenarios and Cognitive Bias. In Risk Metrics for decision making and ORSA. Society of Actuaries. Schaumburg, Illinois. pp. 15-17 http://www.casact.org/pubs/Risk_Essays/orsa-essay-2012-campbell.pdf

4.4 Training/behavioural management

Behavioural economics has many applications for insurance. There is extensive work on typical human biases in reasoning about probability and risk. We know that these biases can become quite acute when people have to make decisions in environments foreign to those our ancestors faced. In particular, insurance forces underwriters to make decisions based on large bodies of somewhat ambiguous data and to handle small probabilities of extreme catastrophe. Because people often make the same kinds of mistakes in these types of situations, there is serious potential for underwriters to introduce systemic risk into the market.

Conversely, there have recently been large-scale experiments that investigate the features of people who are especially good at the task of predicting highly uncertain events[27], finding that there are indeed individual differences. Research in the growth of expertise has shown that it can be trained for some tasks, especially the ones that provide feedback, can be decomposed into parts, and have adequate decision aids[28].

There are two main methods to help mitigate the effects of cognitive bias. First, we can actively train modellers and underwriters to be less susceptible to some of the most important biases. While we know that full elimination is impossible, some successes in other industries give us reason to hope that the right kind of training will be useful.

Second, we can try to set up organisations in a way that will make them more resilient to the effects of behavioural bias even if the participants themselves are biased[29]. For instance, it's often helpful to require participants in a meeting to write down their initial positions before discussion to combat some of the effects of anchoring. Instituting policies of checking randomly selected decisions or model results can counteract asymmetric error checking. But such policies need to be anchored in the organisation: everybody needs to understand why they are done and there must be support from management even when they are inconvenient.

Case studies in other domains

While behavioural biases have not been widely recognized until recently, some areas have begun taking steps to investigate and counter them:

- The oil and gas industry has a long tradition of investigating cognitive bias – since mistakes can be very costly[1]. Anchoring and overconfidence in probability estimates are found, and training and experience have rather weak – but positive - effect in ameliorating them[2].
- Military decision-making has been found vulnerable to bias, and in some quarters training efforts have been attempted[3].
- Intelligence analysis is highly vulnerable to bias[4] and there is evidence that biases on multiple levels can impair national security[5]. The US intelligence community has investigated various debiasing and decision support methods in a realistic setting, as well as structured analysis methods[6].

[1] Krause, T. (2010). High-reliability PERFORMANCE: Cognitive biases undermine decision-making. ISHN, 44(9), 46. http://bstsolutions.com/en/knowledge-resource/163-high-reliability-performance-cognitive-biases-undermine-decision-making
[2] http://www.psychology.adelaide.edu.au/cognition/aml/aml2/welsh_aml2.pdf http://csjarchive.cogsci.rpi.edu/proceedings/2007/docs/p1647.pdf
[3] Janser, M. J. (2007). Cognitive biases in military decision making. ARMY WAR COLL CARLISLE BARRACKS PA. Davis, P. K., Kulick, J., & Egner, M. (2005). Implications of modern decision science for military decision-support systems. Rand Corporation.http://www.au.af.mil/au/awc/awcgate/milreview/williams_bias_mil_d-m.pdf
[4] Heuer, R. J. (1999). Section III: cognitive biases. In Psychology of intelligence analysis. United States Government Printing.
[5] Yetiv, S. A. (2013). National Security Through a Cockeyed Lens: How Cognitive Bias Impacts US Foreign Policy. JHU Press.
[6] Cook, M. B., & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. Human Factors:The Journal of the Human Factors and Ergonomics Society, 50(5), 745-754

[27] http://www.goodjudgmentproject.com/, Tetlock, P. (2005). Expert political judgment: How good is it? How can we know? Princeton University Press.
[28] Shanteau, J. (1992). Competence in experts: The role of task characteristics. Organizational behavior and human decision processes, 53(2), 252-266
[29] Lovallo, D., & Sibony, O. (2010). The case for behavioral strategy. McKinsey Quarterly, 2, 30-43. http://www.mckinsey.com/insights/strategy/the_case_for_behavioral_strategy

4.5 Final words

Systemic risk is everybody's problem: by its nature it is shared. There is often responsibility for many indirectly affected stakeholders who are not themselves involved in the practices that create the risk. Mitigating such broad risks is hence socially and morally significant. Being a good citizen of a community requires us to "clean up" the risks we impose on others.

Systemic risk is often intimately tied to what makes the system useful. We cannot reap the benefits of modelling without risking that our modelling practices sometimes mislead us. But we can avoid overconfident gambling and actually try to measure and manage our systemic risks.

Throughout its long history the insurance industry has specialized in managing risk regardless of what domain the peril exists in: over time new perils – whether airplanes or cyber – emerge, are handled, and eventually become profitable. It may be that meeting the challenge of systemic risk of modelling will be the next step in this sequence. If so, it will be useful far outside the confines of insurance.

Glossary

Word	Definition / description
"What if" scenarios	Changing values inserted into models or their parameters to determine the sensitivity and consequences; sometimes also includes entire scenarios of possible events
1 in 100 TVaR	(Tail Value at Risk) The expected amount of loss given a loss equal or larger to the 1% VaR
1 in 200 VaR	The amount of loss expected to be exceeded in only 0.5% of relevant time (Value at Risk) periods
Average annual losses	The minimum amount of premium to charge to cover losses over time
Back testing	Testing a model on past data or time periods in order to gauge its performance
Basel II	The second Basel Accords, recommendations on banking law and regulation. In particular, it amends international standards on how much capital banks need to hold to guard against financial and operational risks
Capital buffers	The amount of money a financial institution is required to hold in order to avoid excessive insolvency risk
Cyber	Risk from failures of information technology
Delphi	A method for systematically combining the opinions of experts, originally intended for forecasting. The experts answer questionnaires in two or more rounds. After each round they see the joint distribution and motivations, updating their own responses.
Dependent validation	Validation undertaken or coordinated by model users who are involved in the production usage, development, parameterisation, testing and/or operation of the external catastrophe model that feeds the Internal Model. This is opposed to independent validation carried out by validation risk experts e.g. within the Risk department who are removed of the production, development, parameterisation, testing and/or operation of the catastrophe models in the Internal Model.
Economic Scenario Generators	Models generating a scenarios for economic risk drivers such as interest rates, credit risk, inflation, equity returns, real estate returns etc.
Exceedance probability	A diagram showing the estimated probability that losses will be larger than curves different values.
Hybrid systemic risk	Systemic risks that span more than one system, such as insurance and capital markets, or energy and food security
ILS	Insurance Linked Securities
Internal model	A company or institution's model of how different kinds of risk – insurance risk, risk, capital operational risk etc. – will affect it. In insurance, Solvency II requires calculating capital requirements using their internal models.
Linearly weighted	A series of numbers are multiplied by fixed weight numbers and summed. This makes the output depend more on the numbers with large weights.
Lloyd's	Lloyd's of London is an insurance market, a corporate body acting as a quasiregulator of the London insurance market.
Machine learning	Techniques for automatically extracting patterns from data, allowing software to predict, classify or approximate new data.

Word	Definition / description
Metamodel	A model of the modelling process.
Minimum capital requirements	The minimum level of capital required by Solvency II to be held by an insurer. The higher being the SCR (Solvency Capital Requirement) which is the prudent measure.
Model entropy	A measure of how concentrated or dispersed reliance on models is: if numerous independent models are used the model entropy is high, while if most model use is based on a single model it is low.
Model independent analyses	Multiple uniformly spaced scenarios placed within a geographical polygon to scenario identify the area of maximum accumulation.
Non modelled risks	Sources of non-life loss that may arise as a result of catastrophe events, but which is not explicitly covered by a company's use of existing catastrophe models
Nonlinear regression	Statistical modelling of data where a nonlinear curve is used to approximate how input data contributes to output data.
ORSA	Own Risk Solvency Assessment
Overfitting	When a model describes noise and data artefacts rather than the underlying reality. This commonly happens because the model is excessively complex compared to the number of observations it is fitted to.
P&C	Property and Casualty
Portfolio	The book of business of an insurer or reinsurer, in particular all policies held.
Quasiregulator	Institution that performs many of the same regulatory functions as a regulatory body without specific enabling legislation
Realistic Disaster Scenarios	Stress test scenarios based on various disasters maintained by Lloyd's used to test syndicates and the market
Regression fit	Statistical model fitting a simple linear rule to data, often used for forecasting or approximation.
Reverse stress testing	Instead of testing the consequences of a stressful event on a company, one can analyse risk to find the smallest stress that could cause a given bad outcome.
Risk diversification	Reducing overall risk by having a portfolio covering a variety of (hopefully) uncorrelated risks, making the risk of the whole lesser than the sum of the parts
Risk profile	The estimated risk distribution for a portfolio.
Risk-return profile	The pattern of risk estimated for different returns on a portfolio.
Solvency II	EU programme for a harmonized insurance regulatory regime.
Spider bombs	Multiple uniformly spaced circular scenarios placed within a geographical polygon to identify the area of maximum accumulation, most commonly used for terrorism analysis.
Statistical power laws	Probability distributions where the probability of events of size x is proportional to x-a where a>1. Such distributions have large tail risk, and show up in estimates of many catastrophe risks.

Word	Definition / description
Stochastic models	Models where numerous randomly generated events and their consequences are simulated in order to estimate the combined probability distribution of outcomes.
Stress Scenarios / testing	Analysis of how much a given crisis will affect a company, financial instrument Stress or market
Systemic Risk of (SRoM)	Systemic Risk of Modelling. Inadvertent increases in (shared) risk due to use Modelling of risk models.
Tail risk	The risk from extreme events far away from the median events. If the risk probability distribution is heavy-tailed the tail risk from even very rare events can dominate the total risk.
Technical price	A price expected to generate a certain expected loss ratio.
Tragedy of the commons	Situation where individuals act rationally according to their own self-interest, but the end result is against the best interests of the whole group
UK ICAS	UK Individual Capital Adequacy Standards.
Underwriting cycle	Insurance underwriting has a tendency toward cyclicality. First premiums are low due to competition and excess insurance capacity. After a natural disaster or other cause a surge in insurance claims less capitalized insurers are driven out of business. Less competition and capacity leads to higher premiums and better earnings, which attracts more competitors.
Use test	The Solvency II requirement that insurance companies actually use the internal model that generates their estimate for capital buffers.

Appendices are available at

http://www.amlin.com/~/media/Files/A/Amlin-Plc/Systemic_Risk_Scorecard_Appendices

Notes

The Leadenhall Building, 122 Leadenhall Street, London EC3V 4AG
Switchboard +44(0) 207 746 1000
www.amlin.com