



Institute  
and Faculty  
of Actuaries

# Zero to hero

**Presented by**

Isabelle Williams  
Oliver Grossman  
Joanna Thornett

**Contributions by**

Hugo Wood  
Richard Holloway





Institute  
and Faculty  
of Actuaries

**Can we build a capital  
model in an open-  
source framework,  
from scratch, using  
ChatGPT?**



# Introduction

*Is the Lloyd's insurance market ready for open-source capital modelling?*

*George Wright,  
Vounder Analytics*

1. Can an actuary with no capital modelling experience design a capital model using ChatGPT?
2. Can an actuary with very little coding experience learn to code using ChatGPT?



# Why is this something we think is worth trying?

Make IMs more readily available

Reduce specialist knowledge and key person risk

Improve model transparency

Mitigate subscription costs

How easy is it to pick up unfamiliar tasks?

Where can ChatGPT take us?



# The steps involved



Idea



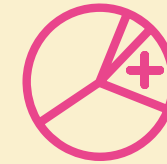
Framework



Data



Model



Validate



Institute  
and Faculty  
of Actuaries



Institute  
and Faculty  
of Actuaries

**Can we build a  
framework for a  
simple capital model  
in 3 days?**



# Let's go!



Is

I am an actuary looking to build a **Solvency II compliant internal capital model** for an insurer based in the **UK**.

The insurer writes **only personal motor and home insurance**. Assume that I have access to all relevant internal data. **What do I do?**



*High level  
Some detail on actual modelling process  
Not enough on how to build one though*



# Getting the information you need

## More background information...

- Is Assume that I have **all the data** that I need
- Is I'm only interested in risks that are **required for the SCR calculation**

## ...or specifying focus points?

- Is In your response, highlight the **differences between motor and household** for all risk types
- Is Note the **statistical distributions and methods** that I need to consider for each risk type and each line of business

## ChatGPT answers...



**Reserve risk:** Calculate the **uncertainty** in the insurer's reserves, for example using **bootstrapping** or other stochastic techniques



**Operational risk:** This can be more difficult to model due to the lack of data, but one approach could be to use an internal loss data model.

## ...but is it enough to implement the model?

**How** do we build a bootstrapping model, and how do we determine the 'uncertainty'?

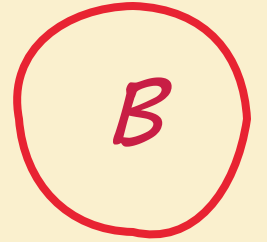
**How** do we simulate **the internal loss data** for the operational risk component and **what model** should we use?







# Reserve Risk



Is

How would I model reserve risk for this book of business? Please highlight any **statistical distributions and techniques** that could be used to model reserve risk.

Is

Could you give **numeric suggestions** as to how the following 5 claim types might be correlated with one another:



**Motor attritional claims and Motor large claims:** There might be a moderate to high positive correlation here. The same underlying factors (e.g., traffic volume, weather conditions) can impact both small and large motor claims. But keep in mind that large claims often have a higher volatility and can be influenced by more idiosyncratic factors (e.g., the specific circumstances of a large accident)



Typically, you would use a **loss reserving method** to estimate your expected reserves and then a **stochastic method to model the uncertainty** around this estimate. There are many loss reserving methods and the choice depends on the available data and the specifics of your business. Here's a broad overview:



# Operational Risk (and simplifications)

C+



Creating a simplified model for operational risk can certainly be a challenge given the diversity of operational risk events.



Instead of using multiple Key Risk Indicators (KRIs), you could select **one representative KRI** that you believe is a good proxy for your overall operational risk

Is

Can you give me **20 examples of operational loss events** that might occur for a UK motor and home insurer?

## Investment mismanagement

Regulatory changes

Fraudulent claims

Supply chain disruption (e.g. repair network)

Employee misconduct

Loss of key personnel

Human error in claims processing

Damage to premises

System outage



# Credit Risk

A-



The expected loss is then calculated as:  $EL = EAD * PD * LGD$ .



Overall, a reasonable LGD could potentially be in the range of 30-50%.



*You need to specify carefully, or it will over-simplify, or go too far, with no questions asked. For example, it suggested overly-complex ways to get transition rates, including using the Merton model and random forests, rather than taking published rates.*



# Generating the data

A-

Is

Assuming I have only UK government bonds in my investment portfolio, what data would I need to calculate the interest rate risk for an internal model approach?



*Gave good examples of the external data you might need  
However, was very generic on **liabilities data***

Is

What internal data would be required in the liabilities data section above?  
What projected cash flows would need to be considered?



*Gave more detail on the data required  
Most required data was also required for other modules*



# What we learned at this stage

- **Great** when looking for high-level information on capital modelling frameworks
- **Great** when using domain knowledge to ask for very specific things – e.g. “what distribution would be appropriate for modelling attritional motor claim frequencies?”
- **Good** at breaking down a framework to add more detail
- **Less good** when asking about (potentially) more subjective expert judgements – e.g. correlations
- **Less good** when trying to check/interrogate the logic behind what ChatGPT has decided
- **Bad** – it never questions one-year vs ultimate





Institute  
and Faculty  
of Actuaries

**How do we build the  
model in open-source  
software?**





# Can we learn to code using ChatGPT?

A+

Jo

How do I read a csv file of data with headings using Python?



Here's an example of doing this using the **pandas** library  
Here's an example using the **csv** library

Jo

I have data called net\_premis and I want to fit a lognormal distribution to this using maximum likelihood estimation in python. How do I do this, in Python?



If you want to fit a lognormal distribution to your data using maximum likelihood estimation (MLE) in Python, you can use the scipy.stats module which provides a suite of functions to fit and work with various probability distributions.

python

Copy code

```
from scipy.stats import lognorm
# MLE estimates
shape, loc, scale = lognorm.fit(net_premis)
```

# Being specific

A+

Jo

In my data set I have one column "nb\_rn" which can take values "NB" or "RN" and one column "wp" which contains the written premium (numerical values).

X and Y are both numerical values.

I want to add a new column to the data set called "aq\_c" which contains the written premium multiplied by X if the "nb\_rn" column contains "NB" and contains the written premium multiplied by Y if the "nb\_rn" column contains "RN".



Sure! You can use the numpy library's **where** function to accomplish this. Here's how you can do it:

```
python Copy code  
  
data['aq_c'] = np.where(  
    data['nb_rn'] == 'NB', data['wp'] * X, data['wp'] * Y)
```







# Debugging and fixing errors

A-

Jo

Okay, I downloaded the data file from SharePoint and then tried to read it in. I'm now getting this error Cell In[6], line 2

```
policy_data_path = 'C:\Users\xxxx\Documents\GIRO talk\Data\policy_data.csv'  
                    ^
```

SyntaxError: (unicode error) 'unicodeescape' codec can't decode bytes in position 2-3: truncated \UXXXXXXXX escape

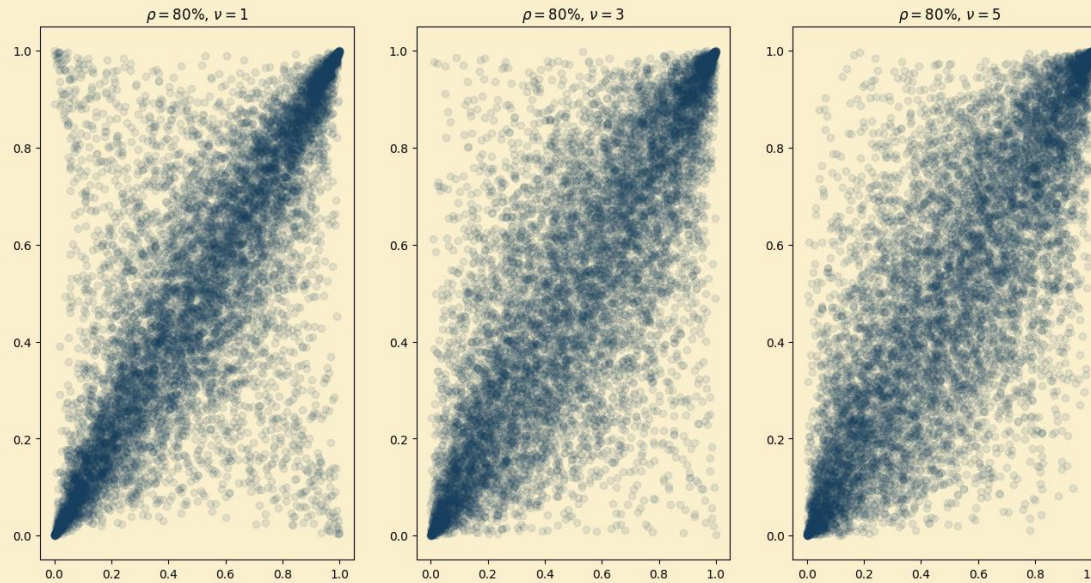
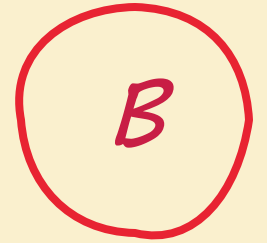


*Double up the backslashes:*  
*Use forward slashes*  
*Use raw strings*





# Ceilings – T-Copulas, an example



- Deciding on the ‘**best**’ package/approach to use.
- Struggled in higher dimensions as complexity increased.
- As good as the literature available on these topics.
- How are parameters typically set? Empirical, Expert judgement, Benchmarking...



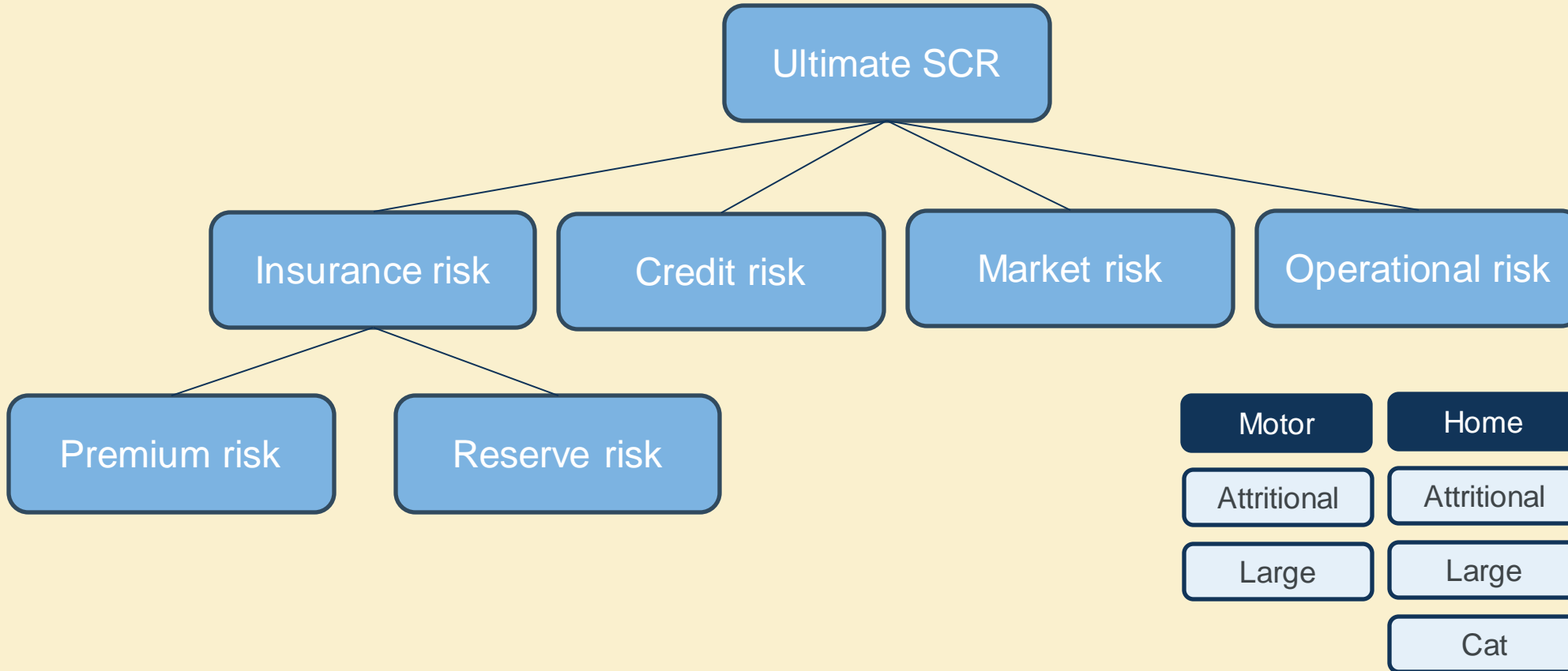


# What we learned at this stage

- **Great** when looking to learn to code using well-documented, commonly used packages – eg pandas, numpy, scipy
- **Great** when looking to learn simple statistical techniques – eg topics that might be covered in an undergraduate university course
- **Good** at explaining the code that it writes and breaking down complex instructions into simple steps
- **Less good** when using more niche packages – eg Chainladder package
- **Less good** when trying to use more complex statistical techniques, or techniques that have varying naming conventions – eg copulas



# Model Outline





# Validation Demo



Institute  
and Faculty  
of Actuaries



Institute  
and Faculty  
of Actuaries

**What can we reveal?**

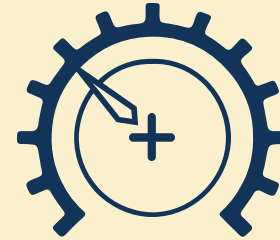


## Strengths



- ✓ **Great** at providing a framework, given a certain level of domain knowledge
- ✓ **Good** at suggesting relevant simplifications, given a certain level of domain knowledge
- ✓ **Great** for learning to code!

## Weaknesses



- × **“Hallucinates”** and will often accept incorrect judgements from the user without question
- × Can make **inappropriate expert judgements** and simplifications
- × Hits an **“academic ceiling”** with some statistical concepts

## Opportunities



- \* **Accelerates** the pace of projects
- \* Enables **better communication** between Data Science & Actuarial
- \* Allows for **more complex, more company-specific** modelling

## Threats



- !! Market disruption to **existing software providers** and **existing capital teams**
- !! **Over-reliance** on ChatGPT’s opinion
- !! Review issues – how do we ensure models are **auditable**?



Alright, **fellow actuary!**

Here's a high-level overview of the key risks to consider for both types of policies and the distinctions between them...

In the context of Solvency II, it is appropriate to **consider Incurred But Not Reported (IBNR) claims** within the **reserve risk module rather than the premium risk module.**

...6. **Liquidity risk...**

Creating a simplified model for operational risk can certainly be a challenge given the diversity of operational risk events...

Home large claims and Home catastrophe claims: **This could have a high correlation.** Catastrophic events can often lead to an increased number of large claims.

...4. **Ignore Scenario Analysis...**

## ChatGPT's Best Gaffes

(In the context of building the market risk component)

Difference: Home insurance might be **more sensitive to property risk** due to direct connections with the housing market.

**Motor Insurance:** The claims development process is usually quicker, and **claims are usually settled within a few years.** This leads to lower uncertainty in the reserves.

(In the context of building the credit risk component of an **internal** capital model)

**Gives standard formula approach for the credit risk module**

