Institute and Faculty of Actuaries

Measuring Model Risk: Real Life Losses, Stress Testing and Fitness for Purpose

Andrew D Smith and Nirav Morjaria
Members of the Model Risk Working Party

04 June 2015

Agenda

1. Introduction
2. Mitigating model risk
   - Sensitivity testing
   - Companion models
   - Driver based scenarios
3. Fitness for purpose
4. What have we learned?

Institute and Faculty of Actuaries

Introduction

04 June 2015

What is model risk?

- **Working definition of a model:** "Any quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates."
- **Why do we need models:** The environments that financial firms have to navigate, and the portfolios and strategies they have to employ, are complex. We need models to help make the right decisions as human intuition and reasoning are not sufficient.
- **Model risk is the risk of adverse consequences from decisions based on incorrect or misused model outputs and reports.** Can lead to financial loss, poor business and strategic decision making, or damage to reputation.
- **Two main causes of model risk:**
  1. Model has **fundamental errors** and produces inaccurate outputs
  2. Model may be **used incorrectly or inappropriately**

Models behaving badly… Some examples

Name	When	What happened	Misstatement and impacts
West Coast Mainline bid	2012	Model used to assess rival bids inconsistent and incorrect conclusion drawn	– Cost to UK taxpayers over £50m – Re-run tender – First Group's business plan damaged
Welsh NHS spending cuts	2011	Spreadsheet calculation error in think-tank's assessment of spending cuts	– Cuts overstated by £130m – Reputational damage
Mouchel Pension Fund	2011	Independent actuaries made an error in a spreadsheet for the scheme valuation	– CEO resigned – Share price fell by a third
AXA Rosenberg	2011	Spreadsheet error over-estimated client investment losses, failed to declare mistake	– $242m fine – Reputational damage
JP Morgan (London Whale)	2012	Ignored control warnings, changed how VaR measured	– $6bn losses – Spreadsheet error at least £250m
Fidelity Magellan	1995	Omission of minus sign lead to over-statement of capital gains and distribution not made	– $2.6bn overstatement – Reputational damage
US Federal Reserve	2010	Spreadsheet error in Fed's Consumer Credit calculations	– $4bn error – Reputational damage
Millennium Bridge	2000	Design calculations and modelling overlooked the fact the Bridge had a resonant frequency of lateral motion similar to that of the gait of pedestrians	– Bridge closed for 18 months – Remedial works cost £5m
LTCM Hedge Fund	1997	Lack of stress testing	– $4.5bn
Over-reliance on Gaussian copulas	2007-12	Mis-price risk of CDOs, poorly understood, over-reliance	– Important role in the 2008 financial crisis

5

Institute and Faculty of Actuaries

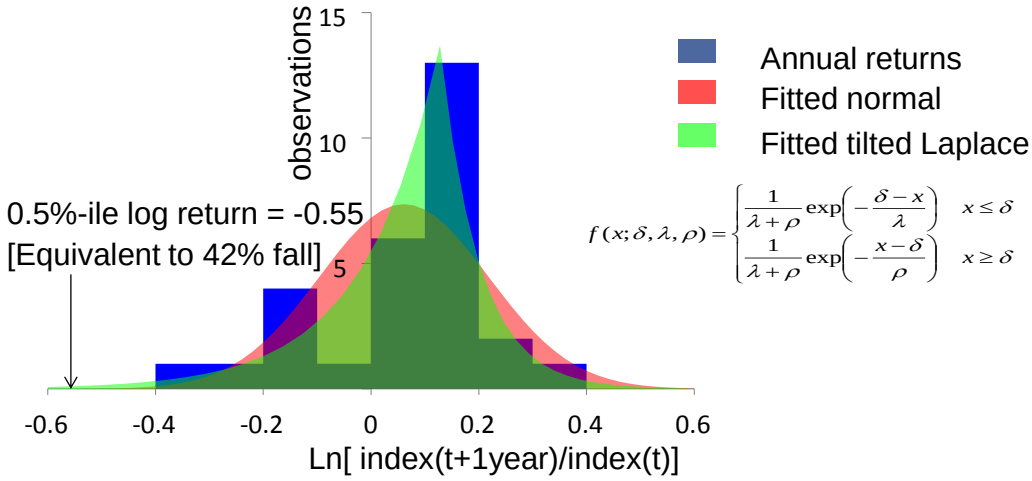
Mitigating Model Risk Idea #1

Sensitivity Testing

04 June 2015

Understanding Model Choice

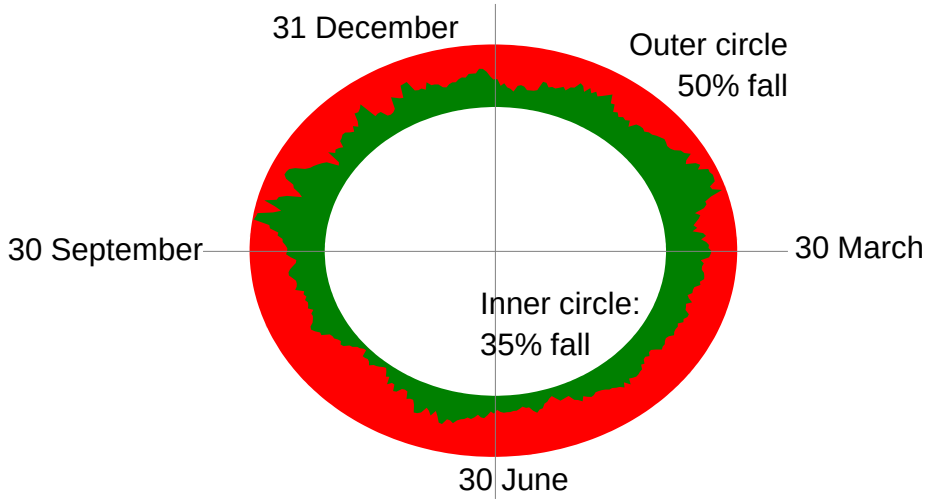
Fitting Distributions to Equity Returns



7

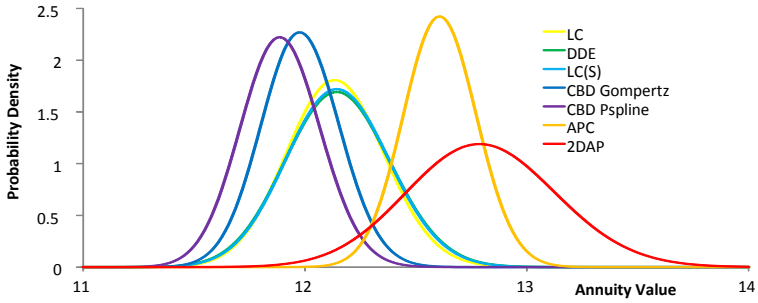
Impact of Starting Point

Choice of date for sampling equity returns



8

Choice of Model Choices
Alternative stochastic longevity models



Probability Distributions for Male Annuity aged 65 (Based on Currie et al, 2013)

9

How does Sensitivity Testing Help?
What does it teach us?

The Calculations
- How model output responds to plausible alternative parameters
- Sensitivity testing does not show that the parameters are right

What to look for
- Robustness; if we mis-estimate the parameter, how much does solvency capital requirement (SCR) move?
- Cliff effects, where a small change in parameter value causes SCR to explode
- Convexity; whichever way parameters move, the SCR goes up.
- Stressed or blended parameters may be an appropriate response

Another Example
- Reinsurance purchase with a limit at the 1-in-200 loss
- Because buying protection beyond that point has no capital impact
- But this is sensitive to knowing what the 1-in-200 event is
- How should management respond to this cliff effect?

10

Institute and Faculty of Actuaries

Mitigating Model Risk Idea #2

Companion Models

04 June 2015

What is a Companion Model?

Checking and Explaining Main Model Output

The Problem

- Internal models can be large and complex
- It is not easy to communicate why the output is reasonable or how it depends on the inputs

What is a companion model?

- Simpler model
- Using well-known input distributions
- And simple dependency structures
- With analytical solution

How to use a companion model in validation

- Approximate check on model results
- Reasonableness of implied parameters
- Highlight important assumptions
- Insight into model drivers
- Some mathematical skill still required to interpret companion output

Companion Model Example
Using Black-Scholes to validate third party ESG

The Situation
- Valuation of a "better of two" option
- Example: maturity guarantee, guaranteed annuity rate, no negative equity on equity release, no negative TB on WP business

How the Full Model Works
- Fixed time horizon T
- Policyholder gets the better of A_T or B_T, where A and B are complicated functions of third party "market consistent" economic scenarios and a cash flow model.
- Present value calculated as

  PVM = E[D * max{A, B}]
- Here, D is the deflator or stochastic discount factor

Calculate present values of the two alternatives:
- Calculate PVA = **E**[D * A]
- Calculate PVB = **E**[D * B]

Assume these are both positive

Black-Scholes formula
- PVM =

$$PVA\Phi\left[\frac{\ln(PVA/PVB)}{\sigma\sqrt{T}}+\frac{\sigma\sqrt{T}}{2}\right]$$
$$+PVB\Phi\left[\frac{\ln(PVB/PVA)}{\sigma\sqrt{T}}+\frac{\sigma\sqrt{T}}{2}\right]$$

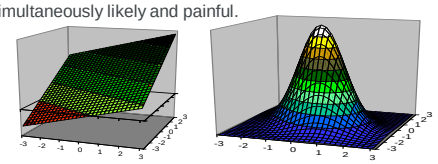
Solve for implied σ
- Assess for reasonableness
- Given asset and liability volatilities

Companion Model Example
Risk Geographies

Introduction
- Combines
  - a mathematical framework for identifying events at given level of confidence
  - an iterative search routine that picks out the most onerous event at the chosen level of confidence
- Scenario based so makes implicit allowance for non-linearity

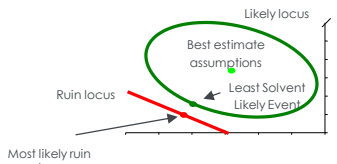
Response and likelihood

We seek to determine combinations of factor values which are simultaneously likely and painful.



Risk geographies process
- Iterative search process
- Convergence generally takes place in 2-5 steps, depending on the number of factors involved and their non-linearity
- Each iteration, requires stress tests to be run for each risk factor and these are fed into a mathematical formula that estimates the "least solvent likely event"

In a picture



Institute and Faculty of Actuaries

Mitigating Model Risk Idea #3

Driver Based Scenarios

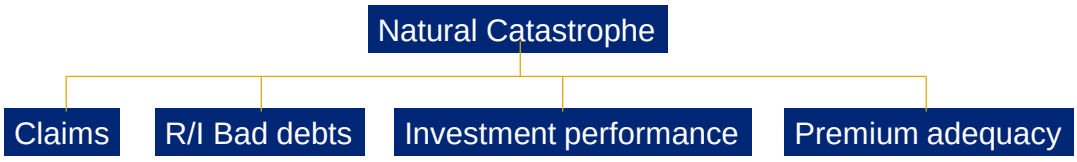
04 June 2015

What is a Driver Based Scenario?

Elicit dependency assumptions by causality not correlation

- Internal models often consider risks on a stand-alone basis and combine with copulas
- Correlations are symmetric: Corr(X,Y) = Corr(Y,X) by definition
- Experts in the underlying risks may advance asymmetric correlations
- Underlying rationale is causal mechanism
- Earthquake might cause a market crash but not the other way round
- For an internal model, joint distributions are important and causation does not matter
- Unless we are misusing what the experts are telling us

Example Driver Based Scenario



Typical Model Structure

Risk driver matrix (100 000 × 50)

Copula component matrix (100 000 × 50)

Net asset matrix (100 000 × 10)
P&L attribution cube (100 000 × 10 × 50)

17

Driver-Based Scenario #1
Emission Limitations

Overview
- There is a familiar story about climate change (global warming)
- What if international efforts to limit emissions are successful?

Scenarios you might consider
- Mortality: winter freeze vs summer heat wave impact
- Global circulation model impact on European storm risk
- What climate scenario underlies current projections?
- Investment impact of unusable fossil fuel deposits, renewable energy sources

18

Driver Based Scenario #2
Ukraine Conflict Escalation

Overview
- Ongoing conflict in Donetsk / Lugansk regions of Eastern Ukraine
- Inconsistent accounts of activities, motivations and supply sources for both sides.

Scenarios you might consider
- Impact of expanded conflict and political instability on Europe's borders
- Possible misjudgements by either side
- Spread of conflicts to Baltic States, re-ignition of Georgia, Moldova, Crimea conflicts
- Disruption of gas pipelines, shipping or aviation
- NATO forces involvement

19

Driver Based Scenario #3
What if UK and Ireland introduced Strict Liability

Overview
- How to determine liability in the event of a transport collision?
- "strict liability" places burden of proof on stronger party
- Eg truck vs cyclist, cyclist vs pedestrian

Scenarios you might consider
- Increased insured liability for motorists who hit cyclists or pedestrians
- Changes in rating factors
- How does this change driver behaviour?
- Impact on modal choice and traffic composition

20

Assessing Stress and Scenario Coverage

- If a risk distribution is wide enough, it should cover plausible stress scenarios
- That is easy to judge in one dimension. For example, we might require that plausible stress scenarios lie within in the fitted range [0.5%-ile, 99.5%-ile]
- In multiple dimensions, if we have a copula (for example T or Gauss) with correlation R) the corresponding test is $x^T R^{-1} x \leq [99.5\%\text{-ile}]^2$, related to the Risk Geographies approach.

Institute and Faculty of Actuaries

Fitness for Purpose

What does this mean?

Doing the Tests That we Know will Pass

Different Kinds of Tests

- Powerful tests must have a reasonable prospect of failure
- Weak tests are unlikely to reject a model
- Power is a function of data size, chosen alternative hypothesis

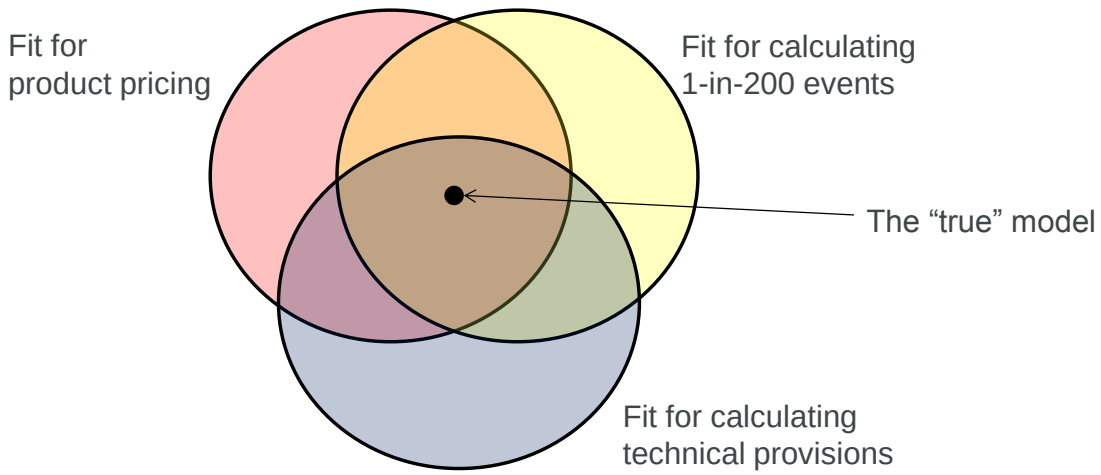
Focus on Cosmetics

- To a naïve reader, a long list of test successes looks more impressive than a shorter list with failures in it
- Question whether readers of validation reports understand the nature of statistical testing
- Insufficient weight to more powerful tests

Example

- Valuing options and guarantees on a with profit fund containing property (a shopping centre)
- Option pricing theory assumes we can take instant long or short positions in 1/1000 of the property without moving the price
- Should validation include test purchases / sales?

Model Fitness for Purpose



What does “Fitness for Purpose” mean?

- Input measures:
  - We have performed the following tests …., which have passed
  - We have followed industry best practice
- Output measures
  - If you use this model for purpose X then nothing can go wrong
  - “The users for whom actuarial information is created can place a high degree of reliance on its relevance, transparency of assumptions, completeness and comprehensibility, including the communication of any uncertainty inherent in the information.” (FRC)

25

Why might you not pick the “true” model?

- Social criteria
  - Parsimony / avoid “unnecessary” parameters
  - Minimise judgemental inputs to ease sign-off process
  - Repeatable calibration and stable outputs
  - Reduce run times by use of analytical formulas
  - Focus effort on tails or middle of distribution
  - If we perform many tests at once (and retain test size) then power of test goes down.

26

Institute and Faculty of Actuaries

What have we Learned?

04 June 2015

Some Conclusions

- Use sensitivity testing to warn when a small parameter change has a large impact on model output.
- Use companion models to test whether model logic functions as intended
- Use driver-based scenarios to check whether a proposed correlation structure captures knock-on effects.
- A model does not need to be perfect, only fit for purpose.
- Please share your ideas on output-based measures of fitness.

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.