Validation of Internal Models

Presented by *Michel M. Dacorogna*

Scientific Advisor to the President of SCOR

ASTIN Colloquium 2016, Lisbon, Portugal, 31st of May to 3rd of June, 2016

Disclaimer

Any views and opinions expressed in this presentation or any material distributed in conjunction with it solely reflect the views of the author and nothing herein is intended to, or should be deemed, to reflect the views or opinions of the employer of the presenter.

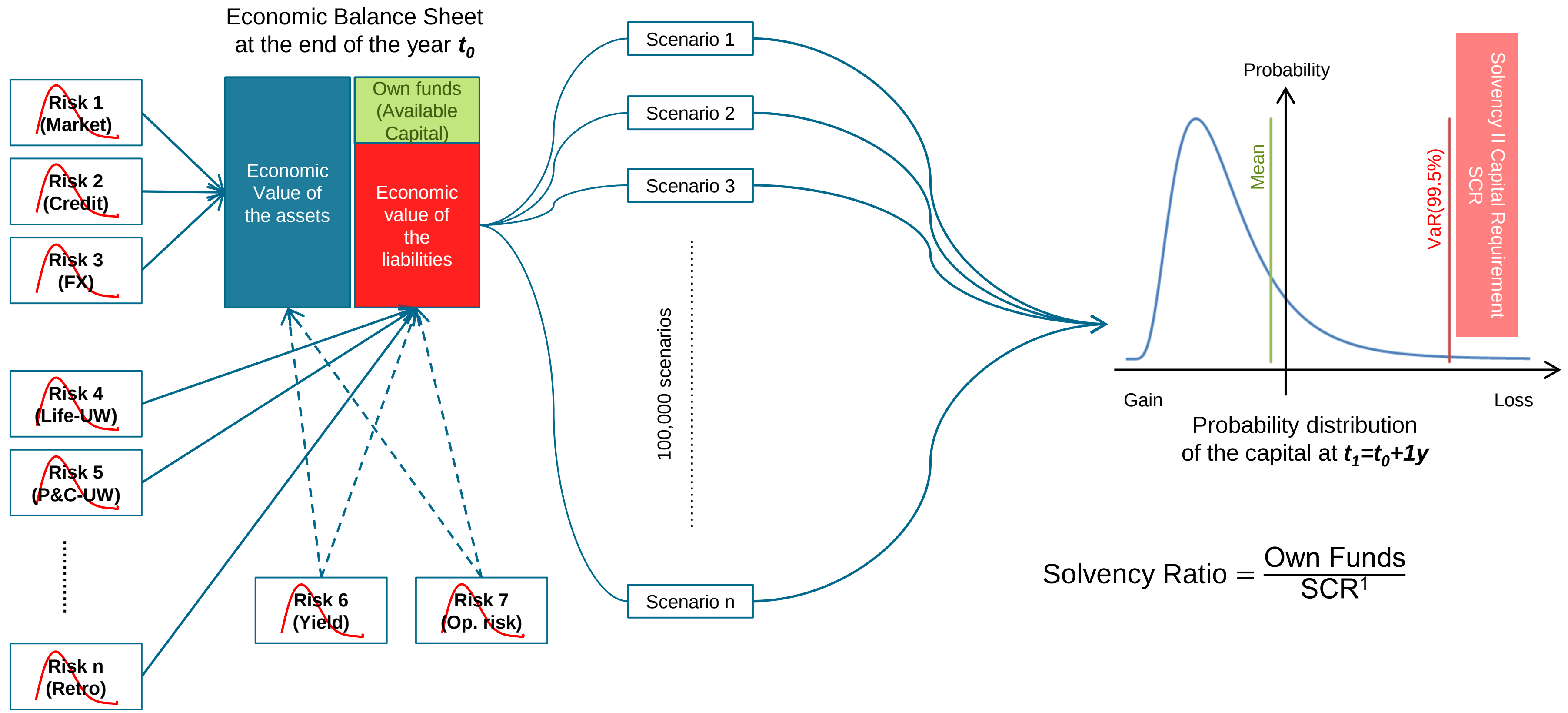
The information, statements, opinions, documents or any other material which is made available to you during this presentation are without any warranty, express or implied, including, but not limited to, warranties of correctness, of completeness, of fitness for any particular purpose.

Agenda

1	Definition and use of internal models
2	Model calibration
3	Testing the various model components
4	Stress testing as a way to check the validity of the model
5	Reverse stress test another way to look at the quality of the model
6	Conclusion

What is an internal model

- An internal model is here to assess *the risk* of the economic balance sheet of the company and to set the required capital



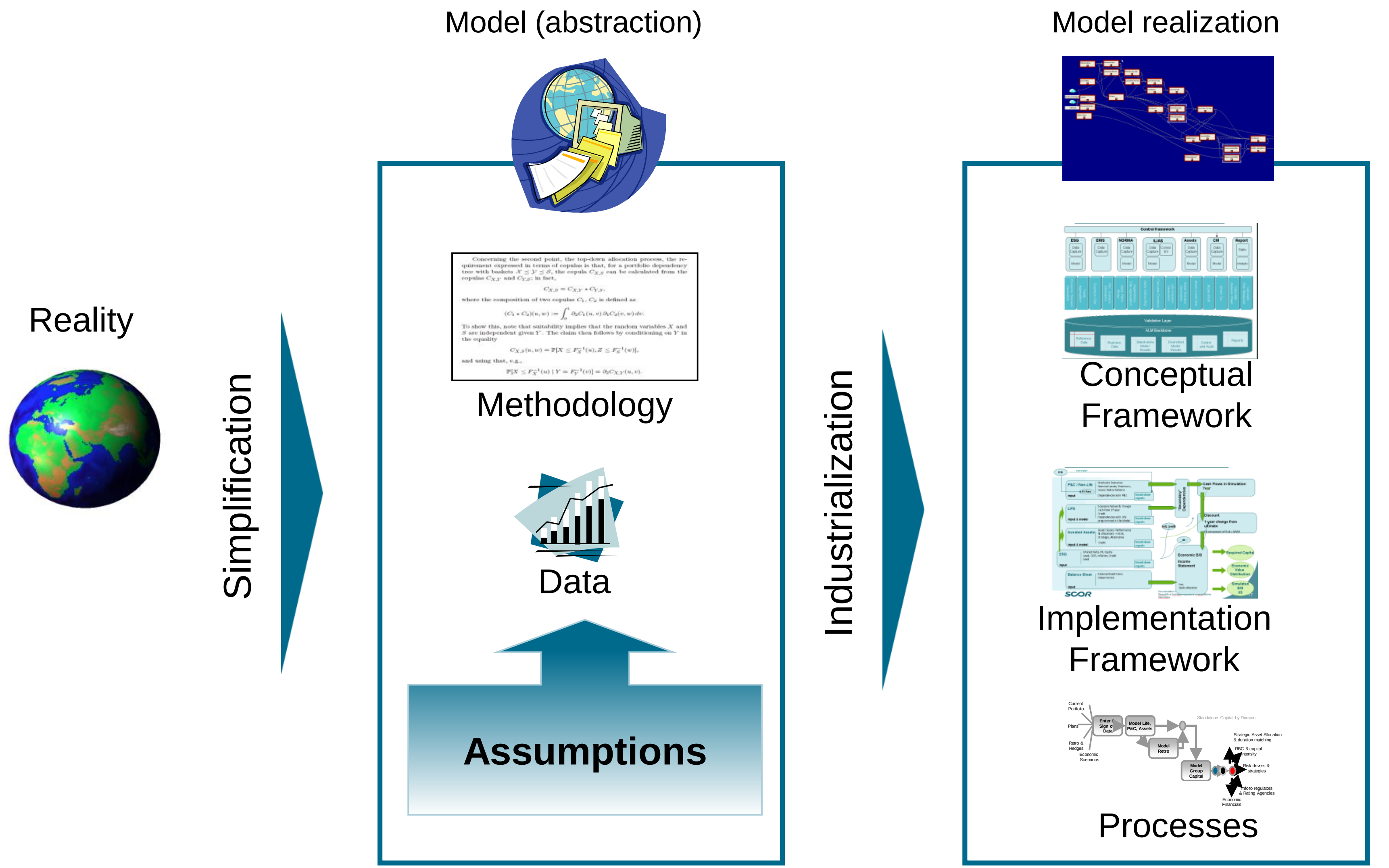
$$\text{Solvency Ratio} = \frac{\text{Own Funds}}{\text{SCR}^1}$$

1) Measured at t_1 but discounted at t_0

Requirements on the internal model

- Internal models should provide a way to assess the *need for capital* to cover the risk assumed
- They should provide a *unified way of communicating about risks* within the company and with outside stakeholders (Solvency requirements, rating agencies, investors)
- They should set the framework for taking *strategic decisions*, balancing risk and return: “Flight Simulator”
- They should allow the *optimisation* of both the asset and liability portfolios by modelling the diversification benefits
- They should make it possible to *measure the economic performance* of the various lines of business

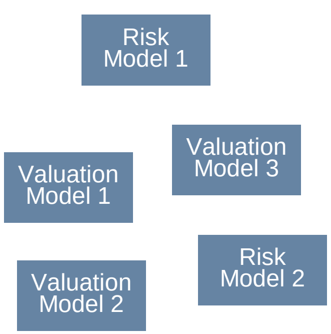
Internal models: development

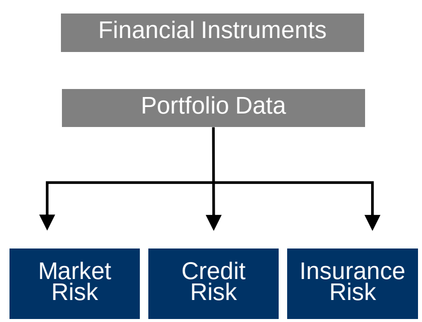


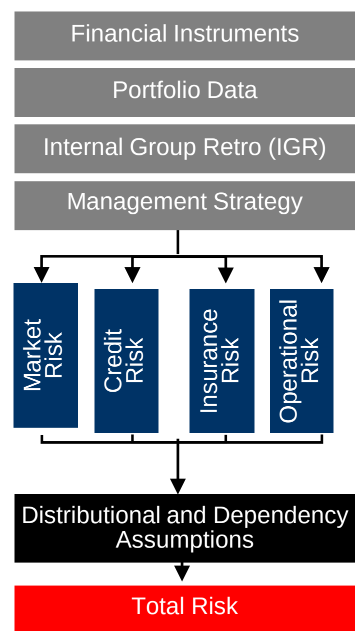
SCOR

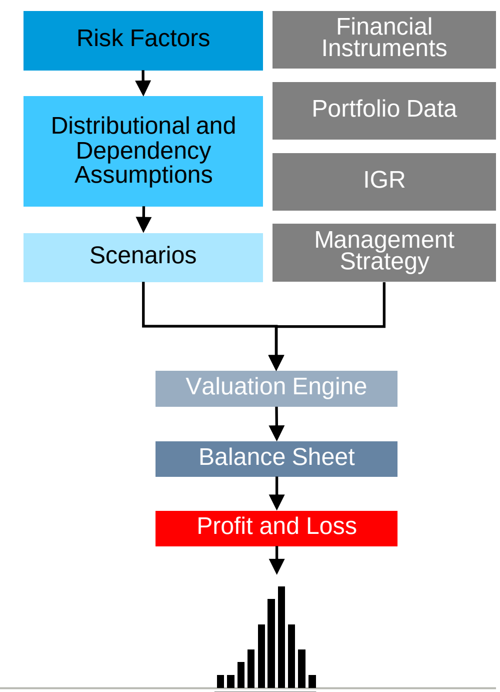
Internal models: historical evolution

Mortality Tables ~1860	De Finetti* 1940	Risk Based Solvency 1995-2000	Capital Management ~2005
	Value Protection	Value Sustainment	Value Creation
Collection of sub models quantifying parts of the risks	Quantification of different risk types with portfolio effects	Risk types are combined to arrive at the company's total risk	Modelling of underlying risk drivers and emphasize on the whole distribution





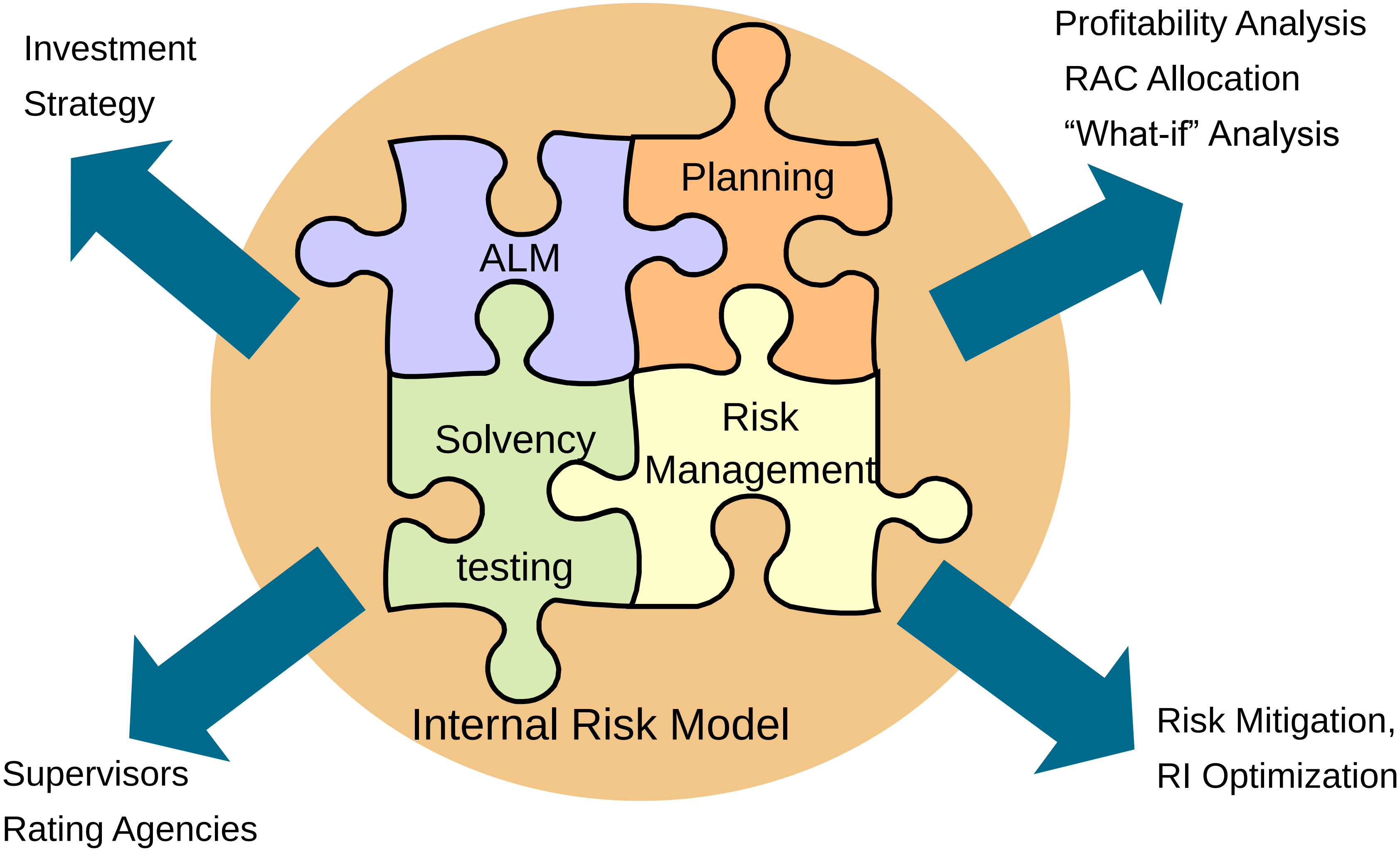




Slide inspired by Philipp Keller

*) B. de Finetti `Il problema dei pieni `, Giornale dell'Istituto Italiano degli Attuari, Vol. 11, No. 1, pp. 1-88, 1940

Internal risk models: applications and benefits



Agenda

Calibration is the first step towards a good model

- Any model needs to *determine few parameters*. These parameters are set looking at data of the underlying process and fitting them to these data
- The pricing and reserving actuaries develop their model based on *statistical tests* on claims data
- The model is composed of *probabilistic models* for the various risk drivers but also to model for the *dependence* between those risks
- Both components need to be calibrated. The most difficult part being to find the right dependence between risks because this requires lots of data, particularly when there is only dependence in the tails
- The probabilistic models are usually calibrated with *claims data* for the liabilities and with *market data* for the assets, or with *stochastic models* like natural catastrophes, pandemic or credit models

How to Calibrate Dependences?

Dependences can hardly be described by one number such as a linear correlation coefficient

We generally use ***copulas*** to model dependences

In insurance, there is often not *enough liability data* to estimate the copulas

Nevertheless, copulas can be used to *translate an expert opinion* about dependences in the portfolio into a model of dependence:

- Select a copula with an appropriate shape
  - *increased dependences in the tail*
    - this feature is observable in historic insurance loss data
- Try to estimate conditional probabilities by asking questions such as “What about risk *Y* if risk *X* turned very bad?”
  - Think about *adverse scenarios* in the portfolio
  - Look at *causal relations* between risks

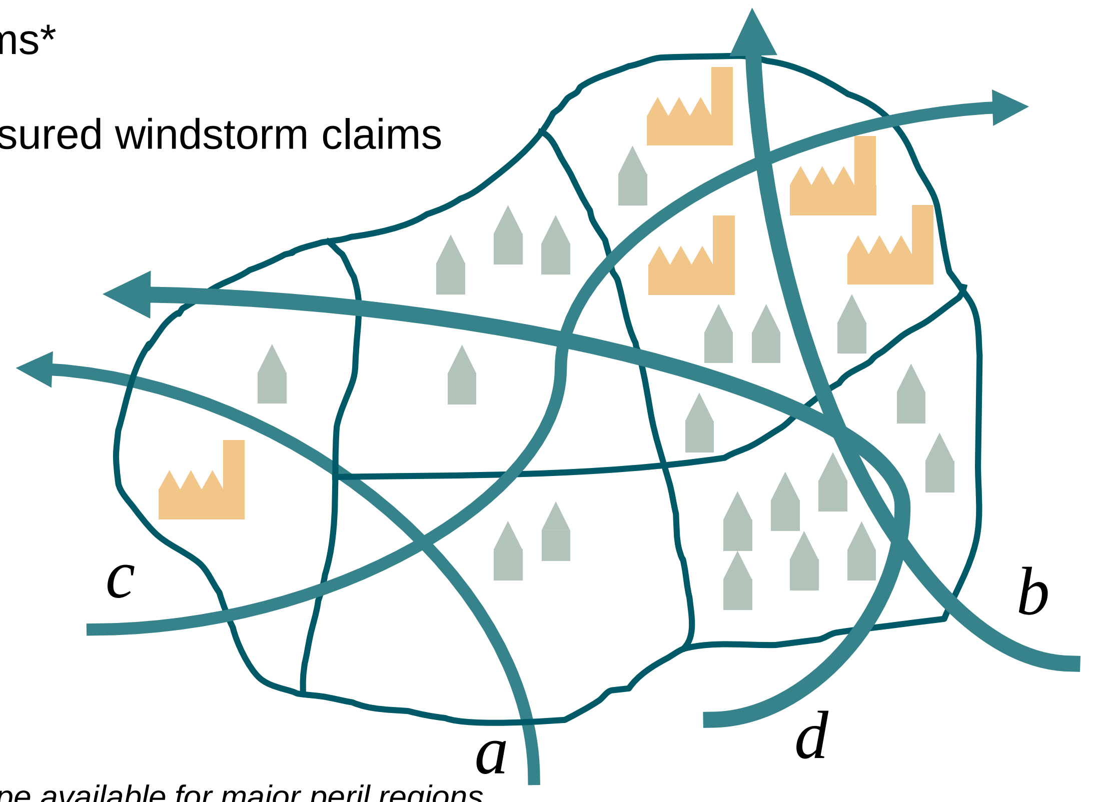
Example: Windstorm

Collect the exposures from all policies per zip-code area in an accumulation control system

- Here: Private homes and industrial plants
- Scenarios = Windstorms*
- Random Variable = insured windstorm claims

Stochastic Simulation

Scenario	Insured Loss
a	3
b	27
c	11
d	8



**There are commercial models of this type available for major peril regions.*

Describing Dependences

Scenario based simulation

- Dependences between random variables modeled on the same scenarios is incorporated automatically
- Example: dependence in "our" windstorm model between losses on *industrial risks* and on *private home owners*
- Building a realistic model of that type is challenging

Distribution based simulation

- Via *joint* simulations of the individual distribution
- Dependent sampling of the joint uniform random numbers

  ➔ *copula*

- Calibration is an issue

Strategy for modeling dependences

- Using the knowledge of the underlying business to aggregate multiple risks, develop a *hierarchical model for dependences* in order to reduce the parameter space and describe more accurately the main sources of dependent behavior

- Once we have determined the structure of dependence for each node there are two possibilities:
  1. If we know a *causal dependence*, we model it *explicitly*
  2. Otherwise, we systematically use *non-symmetric copulas* (ex. Clayton copula) in presence of *tail dependence*

- To calibrate the various nodes, we have again two possibilities:
  1. If there is enough data, we calibrate statistically the parameters
  2. In absence of data, we use *stress scenarios and expert opinion* to estimate conditional probabilities

- For the purpose of *eliciting expert opinion* (on common risk drivers, conditional probabilities, bucketing…), we have developed a Bayesian method combining various sources of information in the estimation: PrObEx

PrObEx: combining three sources of information

- **PrObEx*** is a new methodology developed to ensure the *prudent calibration of dependencies* within and between different insurance risks.

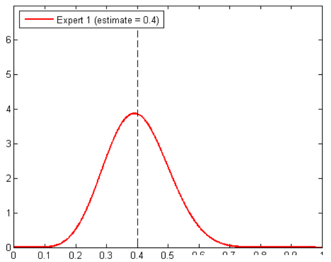
- PrObEx is based on a *Bayesian model* that allows to *combine* up to three *sources of information*:
  - **Pr**ior information (i.e. indications from regulators or previous studies)
  - **Ob**servations (i.e. the available data)
  - **Ex**perts' opinions (i.e. the knowledge of the experts)

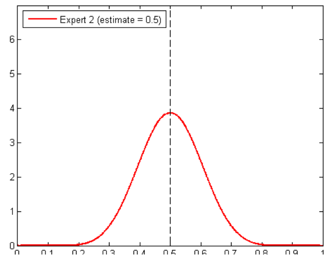
We invite experts to a **Workshop** where they are asked to assess dependencies within their Line of Business.

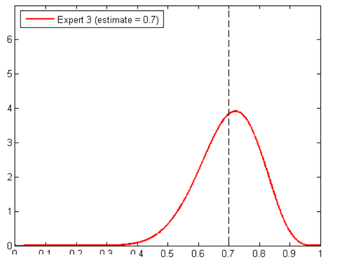
*) P. Arbenz and D. Canestraro, 2012: Estimating Copulas for Insurance from Scarce Observations, Expert Opinion and Prior Information: A Bayesian Approach, ASTIN Bulletin, vol. **42**(1), pp 271-290

PrObEx: combining expert judgements

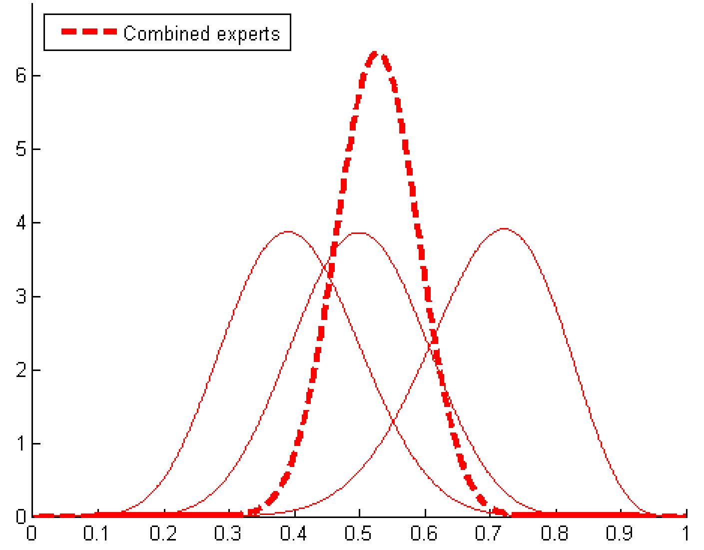
Example: three experts estimate $P\left[X > VaR_{0.99}(X) \mid Y > VaR_{0.99}(Y)\right]$





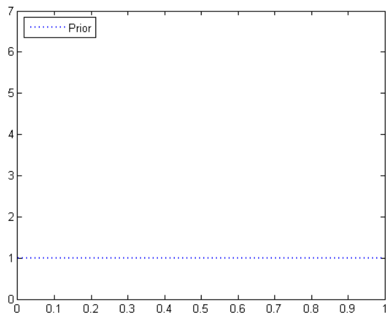


Given these three judgements, PrObEx combines them into a unique, more accurate, estimate

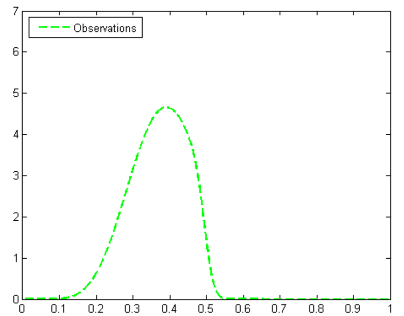


PrObEx: combining up to three sources of information

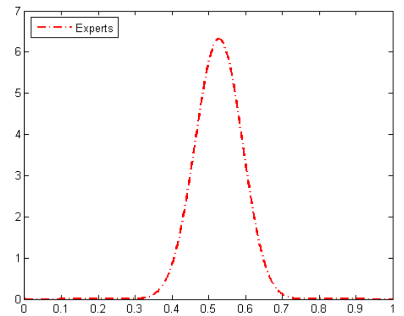
Prior Information



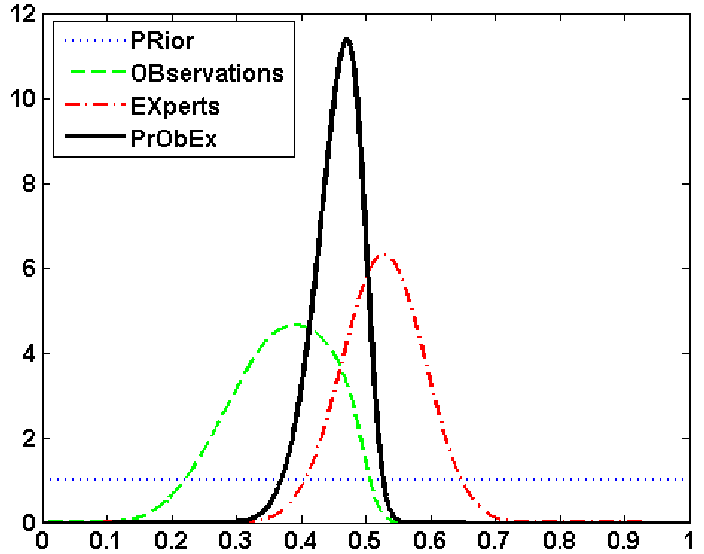
Observation



Expert judgements



PrObEx combines the three sources to provide SCOR with the finest estimate for dependence parameters



Agenda

The important components of internal models

- Every internal model contains *important components* that will condition the results:
  - An economic scenario generator
  - A model for the uncertainty of P&C reserving triangles
  - A model for natural catastrophes
  - A model for pandemic (if there is a life book)
  - A model for credit risk
  - A model for operational risk
  - A model for risk aggregation (dependence)
- Each of these components can be *tested independently*, to check the validity of the methods employed
- These tests vary from one component to the other. Each requires its *own approach for testing*

Testing the quality of ESG scenarios (1/2)

- The ESG produces many scenarios, i.e. many different “forecast” values.
- Thousands of scenarios together define *forecast distributions*.
- *Back testing*: How well did known variable values fit into their prior forecast distributions?
- Testing Method: *Probability Integral Transform* (PIT), (Diebold *et al*. 1998, 1999). Determine the cumulative probability of a real variable value, given its prior forecast distribution.
- We need to define two samples for this:
  - an *in-sample period* for building the bootstrap method with its innovation vectors and parameter calibrations (e.g. GARCH parameters).
  - An *out-of-sample period* starts at the end of the in-sample period and is used to test the generated distributions.

Diebold F. X., Gunther T., and Tay A., 1998, Evaluating density forecasts with applications to financial risk management, International Economic Review, vol. 39(4), pages 863-883.

Diebold, F.X., Hahn, J. and Tay, A., 1999, Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange**,** Review of Economics and Statistics, vol. 81, page 661-673.

Testing the quality of ESG scenarios (2/2)

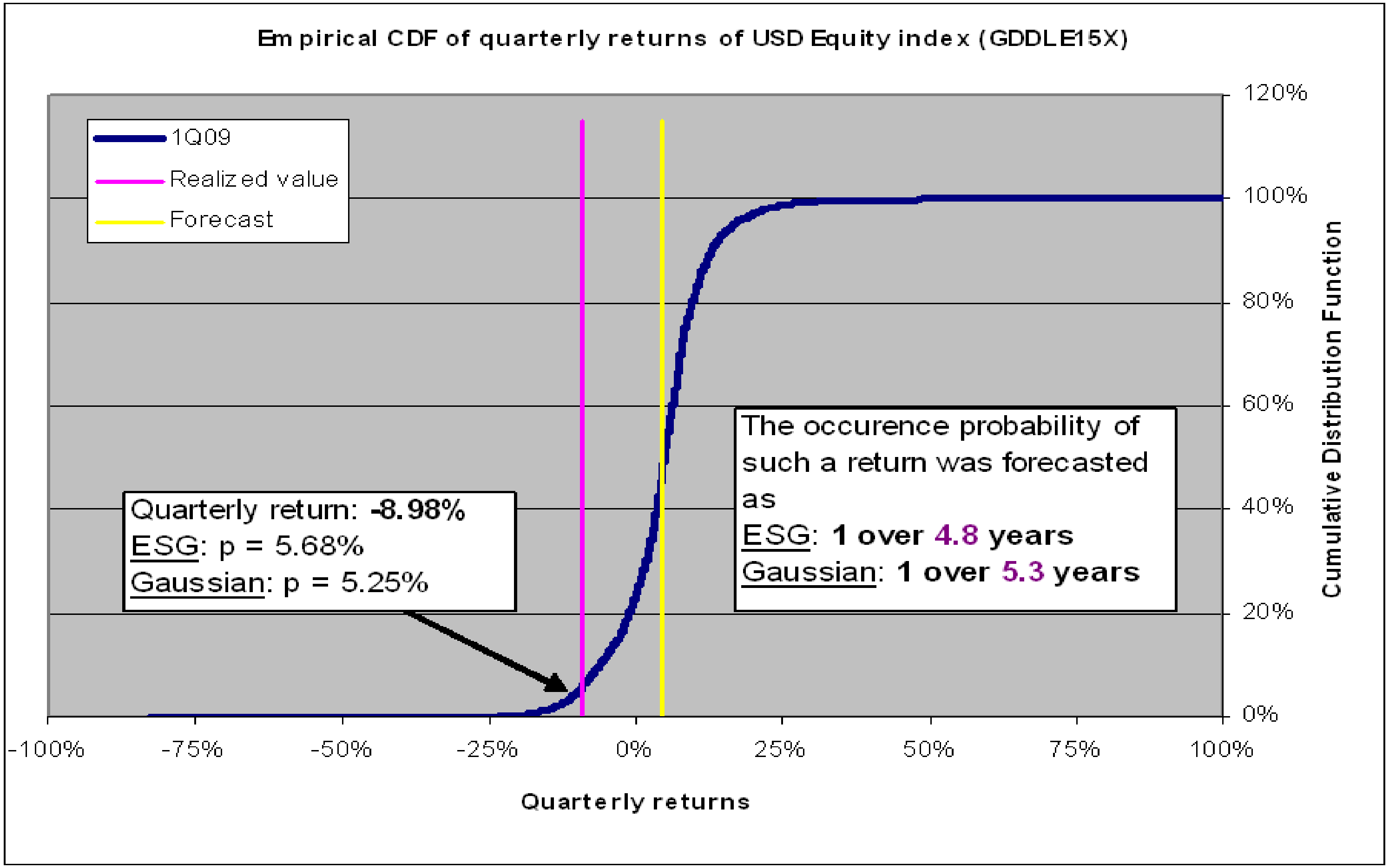
The PIT method is used as follows:

- The scenario forecasts of a variable x at time t_i, sorted in ascending order, constitute an empirical distribution forecast, $\Phi_i(x)$.
- For a set of out-of-sample time points, t_i, we now have a distribution forecast, $\Phi_i(x)$, as well as a historically observed value, x_i.
- The cumulative distribution $\Phi_i(x)$ is then used for the following PIT: $Z_i = \Phi_i(x_i)$.
- A proposition proved by Diebold *et al.* 1998* states that the Z_i are i.i.d. with a uniform distribution $U(0,1)$ if the conditional distribution forecast $\Phi_i(x)$ coincides with the true process by which the historical data have been generated.
- If the series Z_i significantly deviates from either the $U(0,1)$ distribution or the i.i.d property, the model does not pass the out-of-sample test.

*) **Diebold F., Gunther T., and Tay A.**, 1998, *Evaluating density forecasts with applications to financial risk management*, International Economic Review, **39**(4), 863–883.

The ESG scenarios withstood the test of the financial crisis of 2008

Example: Cumulative distribution computed in 30.06.2007 for 31.03.2009



The one year change of P&C reserving triangles

- Modelling the uncertainty of P&C reserving triangles is an important component of internal models
- Testing the quality of the model to compute the one year change is also part of validating a model
- One way of doing it, is to *design stochastic models* to reach the ultimate that can then be used to *test the methods*
- We have done this with simple stochastic models for reaching the ultimate* that allow for *explicit formulae*:
  1. An additive model
  2. A multiplicative model

*) M. Dacorogna, A. Ferriero and D. Krief, 2015, Taking the one-year change from another angle, preprint submitted for publication

Testing the one year change (1/2)

- The additive model is not suited for the Merz-Wüthrich method:

Method	Mean	Std. dev.	MAD	MRAD
Benchmark	18.37	3.92	--	--
COT*, no jumps	19.08	3.93	0.71	4.14%
COT, jumps	18.81	3.86	0.43	2.47%
Merz-Wüthrich	252.89	149.6	234.5	1'365%

- The "mean" in the table is the capital standalone
- The reserves in this model are 101.87
- Capital intensity typical of the Standard Formula

*) The Capital over Time Method has been developed at SCOR by A. Ferriero *Solvency capital estimation, reserving cycle and ultimate risk*, 2016, IME

Testing the one year change (2/2)

- The multiplicative model is better suited for chain ladder and Merz-Wüthrich

Method	Mean	Std. dev.	MAD	MRAD
Benchmark	29.36	21.97	--	--
COT, no jumps	26.75	19.84	2.54	8.19%
COT, jumps	28.30	20.98	1.07	3.48%
Merz-Wüthrich	22.82	15.77	12.7	43.2%

- The results show that all the models underestimate the capital

Testing the dependence model: SCR depends crucially on the right dependence model

- Using the wrong dependence model will lead to either an *underestimation of the SCR* (by neglecting the dependence in the tails) or *an overestimation of the SCR* (by fitting a correlation to a tail dependence as the Standard Formula does)

- We tested this by comparing statistics stemming from a 16-leaves full binary tree, when switching from lognormal(0,1) marginals and *Flipped Clayton copulas* with parameter $\vartheta = 1.36$, to *Gaussian copulas* calibrated either all in *the extreme* (same Quantile Exceedance Probability at 99,5%: "tail correlation") or on the *whole linear dependence* (same "Spearman correlation" coefficient 0.57)

Calibration	Capital Ratio* Gauss/Clayton
Pearson correlation	0.64
Tail correlation	1.06

*) We compute the ratio of the $VaR_G(99.5\%)$ / $VaR_C(99.5\%)$

Testing the Convergence of Monte Carlo Simulations

- We have developed a method to obtain explicit formulae for aggregated *Pareto* distributed risks linked by *Clayton copula**

- We use the results to test the convergence of the Monte Carlo simulations as a function of the parameters

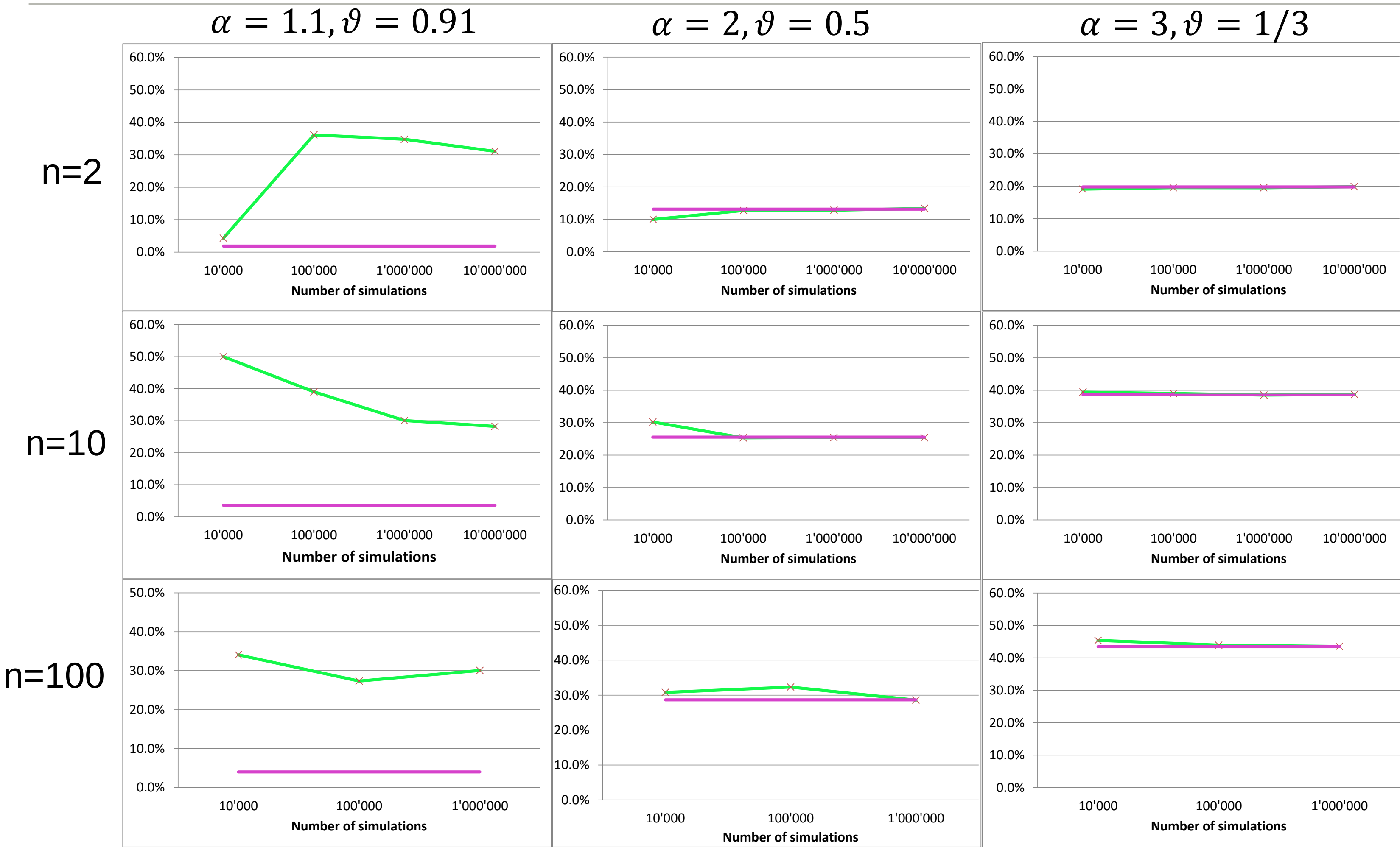
- We compute both the TVaR for the aggregated risks and the diversification benefit of n dependent risks X_i:

$$D = 1 - \frac{\rho\left(\sum_{i=1}^{n} X_i\right)}{\sum_{i=1}^{n} \rho(X_i)}$$

- We see that when the tail is very heavy the simulations do not really converge

*) M. Dacorogna, L. El Bahtouri, M. Kratz, 2016, *Explicit diversification benefit for dependent risks,* SCOR Paper no. 38

Convergence of Diversification Benefit



- The convergence is very good for $\alpha = 2$ and 3 and it does not converge for α =1.1

Analytical values

Agenda

Is it possible to statistically test internal models?

- RAC is computed for a probability of 1% or 0.5%, which represents a 1/100 or 1/200 years event
- In most of the insured risks, such an event *has never been observed* or has been observed only once
- This means that the tails of the distributions *have to be inferred* from data from the last 10 to 30 years in the best cases
- The 1/100 years RAC is thus based on a *theoretical estimate* of the shock size
- It is considered more as the *rule of the game* than as a realistic risk cover
- It is a *compromise* between pure betting and not doing anything because we cannot statistically estimate it

Stress testing the models is crucial

- Testing the output of internal models is thus a must to gain confidence in its results and to understand its limitations

- We just saw that it is difficult, or even impossible, to *statistically test* the model. We can only *stress test* it

- There are at least four ways of stress testing the models:

  1. Test the sensitivity to parameters (sensitivity analysis)
  2. Test the predictions against real outcomes (historical test, via P&L attribution for lines of business (LoB) and assets)
  3. Test the model against scenarios
  4. Study the reasonableness of the extreme scenarios of the Monte-Carlo simulations (*reverse stress-test*)

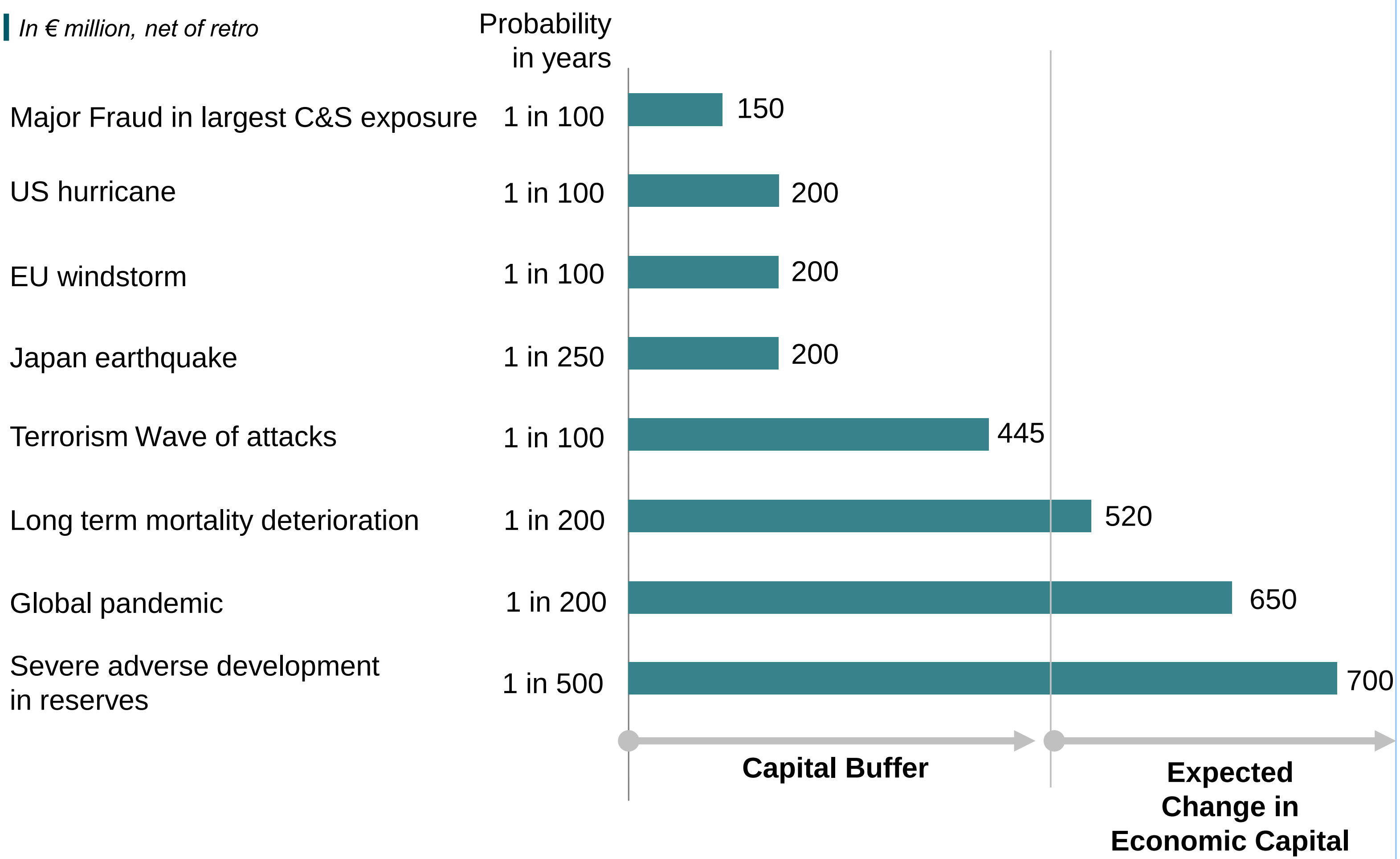
Testing stochastic models with scenarios

- Scenarios can be seen as *thought experiments* about possible future world situations
- Scenarios are different from sensitivity analysis where the impact of a (small) change to a single variable is evaluated
- *Scenario results* can be *compared to simulation results* in order to assess the probability of the scenarios in question
- By comparing the probability of the scenario given by the internal model to the expected frequency of such a scenario, we can assess whether the internal model is realistic and has really taken into account enough dependencies between risks
- By studying the extreme outcomes of the Monte-Carlo simulations, it is possible to determine their plausibility

Capital Buffer to absorb single worst case scenarios

Buffer capital checked against single worst-case scenarios (examples)

In € million, net of retro



Agenda

Making full use of the Monte Carlo simulations

- Stochastic models produce many simulations at each run. These outputs can be put at use to understand the way the model works

- We select the *worst cases* and look at what are the scenarios that make the company bankrupted. Two questions to ask on these scenarios:
  1. Is this scenario *credible* given the company portfolio?
  2. Are there other possible scenarios that do not appear in the worst Monte Carlo simulations?

- This is typically the kind of *reverse back testing* that can be done on the simulations

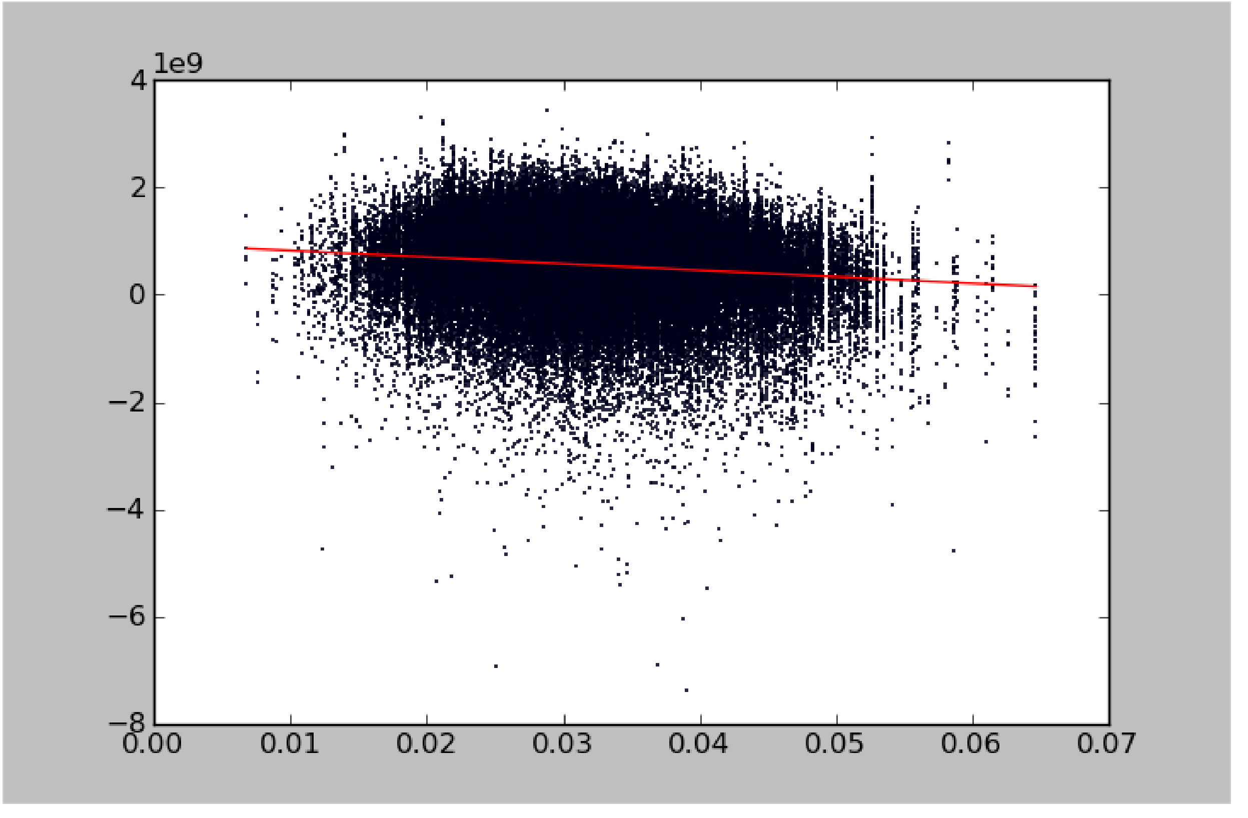
- Other tests are also interesting like *looking at conditional statistics*. A typical question would for instance be: how is the capital going to behave if interest rises?

Reverse Stress Test: Testing the Output of Internal Models

- Internal models generate a huge quantity of data. Usually *little of these data is used*: some averages for computing capital and some expectations
- Exploring the *dependence of results* to certain important variables is a very good way to test the reasonableness of the model
- In the next few slides, we present regression plots, which show the dependency between interest rates and change in economic value (of certain LoB's)
- The plots are based on the full 100'000 scenario's of the Group Internal Model (GIM)
- By analyzing, the GIM Results on this level, we can follow up on a lot of effects and test if they make sense

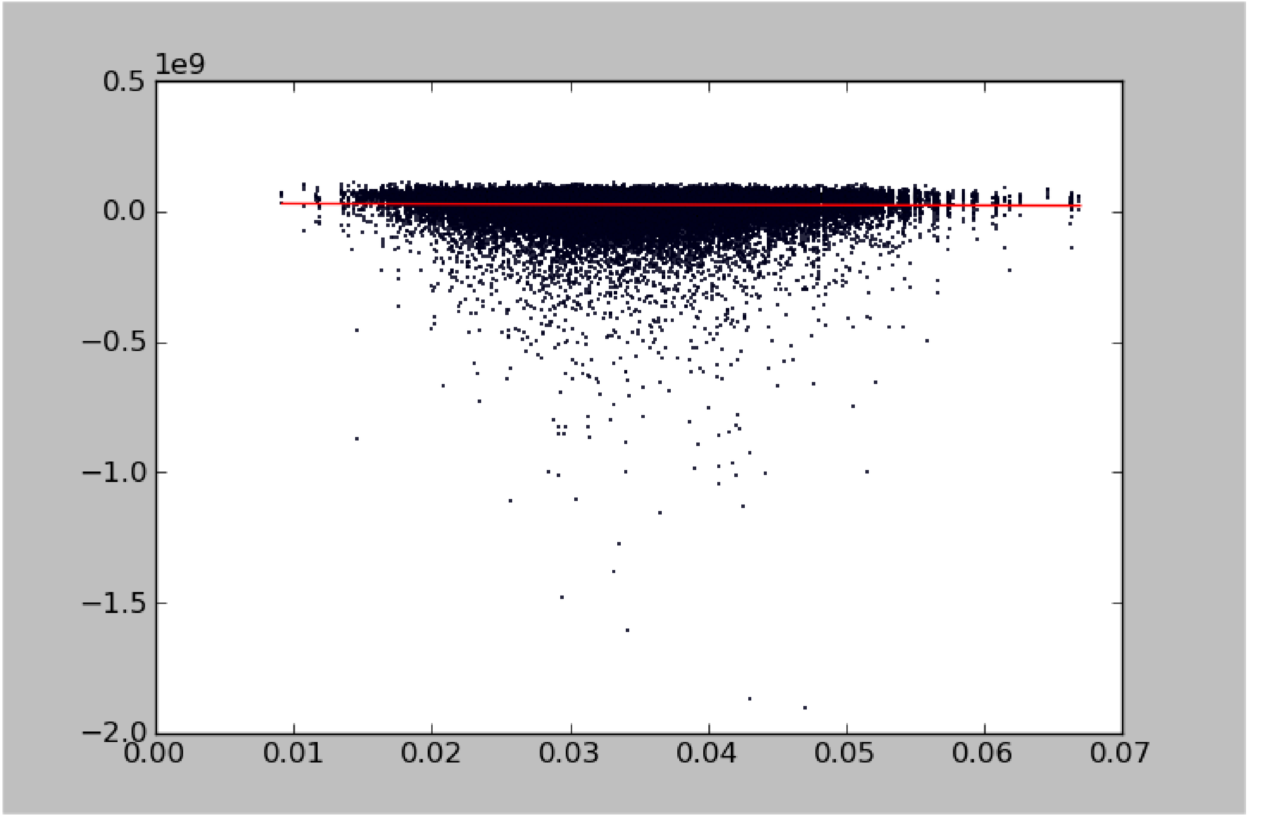
Change of Company Value versus the 4Y EUR Gov.

- 4Y is the typical duration of the P&C portfolio



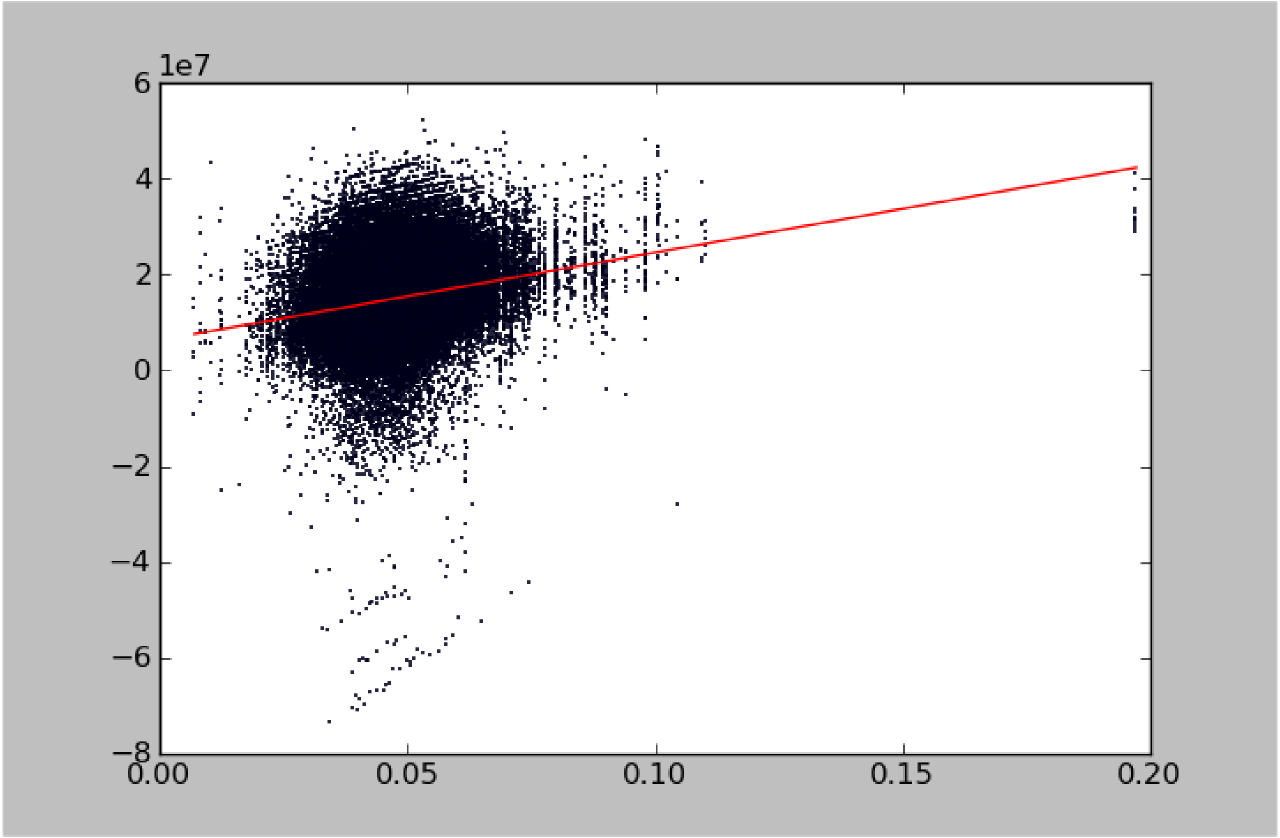
- As interest rate grows the Value of the company slightly decreases decreases (due to an increase in inflation linked to IR increase)

Motor Business versus 5Y EUR Gov. Bond Yield



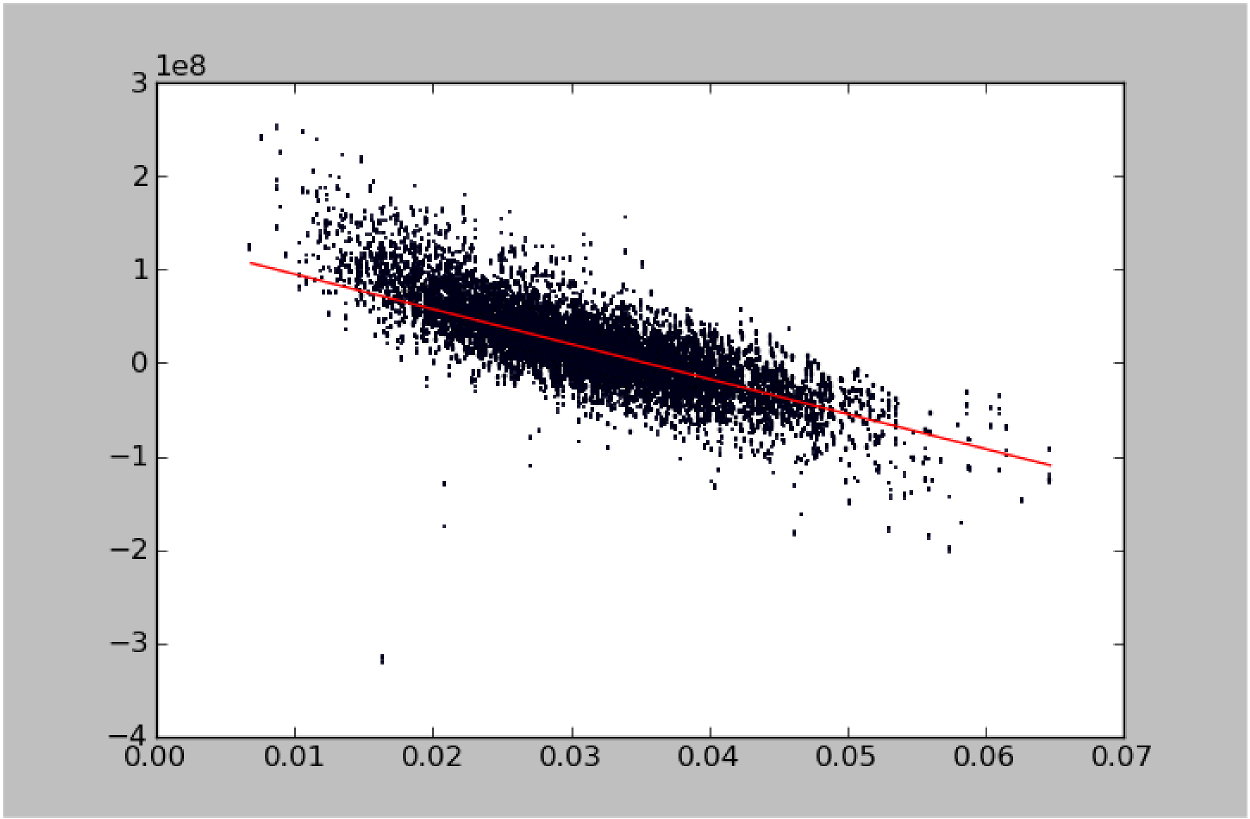
- The value of motor business depends only very weakly on interest rate as it is relatively short tail

Professional Liability (long tail) versus 5Y GBP



- The value of professional liability business depends heavily on interest rate as it takes a long time to develop to ultimate and the reserve can earn interest for a longer time

Gov. Bond Assets versus 4Y EUR Gov. Bond Yield



- Bond value depends mechanically on interest rate. When interest rate increases the value decreases

Agenda

Conclusions (1/2)

- The development of risk models helps to *improve risk awareness* and anchors risk management and governance deeper in industry practices
- Risk models provide valuable assessments, especially in relative terms, as well as *guidance in business decisions*
- It is thus essential to *ensure* that the results of the model delivers a *good description of reality*
- Model validation is the way to *gain confidence* in the model
- It is however difficult because there is *no straightforward way* of testing the output of a model

Conclusion (2/2)

- Validating a risk model requires the use of various strategies:
  - Ensure a *good calibration* of the model through various statistical techniques
  - Use data to *test statistically certain parts* of the model (like the computation of the risk measure, or some particular model like ESG or Reserving Risk)
  - Test the P&L attribution to LoB's against real outcome
  - Test the *sensitivity* of the model to *crucial parameters*
  - Compare the model output to *stress scenarios*
  - Compare the *real outcome* to its *predicted probability* by the model
  - Examine the simulation output to check the quality of the bankruptcy scenarios (*reverse backtest*)