MANAGING EXTREMES

Willis Re

ECONOMIC CAPITAL MODEL VALIDATION

Markus Stricker
Willis Economic Capital Forum

David Simmons
Willis Re

Dave Ingram
Willis Re

Contents

Abstract

The paper considers Economic Capital Model (ECM) validation from different viewpoints: regulators, management, rating agencies and parallel developments in the banking sector. Validation is a necessity, and yet standardized ECM validation processes are underdeveloped. This can lead to inefficient and person-dependent validations, which are certainly undesirable for both management and regulators. We outline explicit guidance for ECM validation to foster a more standardized, efficient process. Those involved in a model validation project are our intended audience; model developers may also benefit from the paper, but will have to translate the validation guidelines into development guidelines.

Many articles on ECM validation formulate rather general principles that are not easily translated into explicit guidance. We believe this is a consequence of imprecise definitions of model risk. Since the purpose of validation is to assess the level of model risk, it is imperative to work with a practical definition of model risk. In this paper we provide a definition of model risk consisting of five sub-categories: conceptual risk, implementation risk, input risk, output risk, and reporting risk. Our validation guidance in the following sections is derived from this definition.

> *MODEL RISK*
>
> *A sound process for economic capital model validation requires a clear definition of model risk. In this paper we identify five sub-categories: conceptual risk, implementation risk, input risk, output risk, and reporting risk.*

Before providing detailed assessment guidance for each of the five model risk sub-categories, we propose a classification scheme for model validation results. The embedding into a process yields a natural sequence for assessing the model risks.

The discussion of conceptual risk starts by emphasizing the need for careful description of the purpose of the model, its applications, and its users. Without these it is impossible for the validation team to assess the adequacy of the concepts and whether the presentation of the output is useful decision support for the users. Clear documentation of the limitations of the model concepts is very important.

Most of the guidance in the implementation risk section comes from best practices for software engineering. For complex software, the realistic question is not whether it contains errors, but rather whether the errors it contains are substantial. Various test techniques are outlined.

The input risk section discusses the different types of input and how they can be validated or benchmarked. This category of risk is typically the most familiar to actuaries, although we feel that there is often too much emphasis on extracting all the information from internal data and not enough attention to peer group benchmarks.

Only if conceptual, implementation and input risk have been assessed positively does it make sense to assess output risk. We make a distinction between output risk and reporting risk: the former deals with the full data set of outputs and checks whether the model yields reasonable values, while the latter deals with the manner in which selected outputs are presented to users. The outputs must be calculated correctly, but this alone is not sufficient; even correct outputs may be highly sensitive to input parameters, which is an undesirable model feature.

Reporting risk deals with the communication of model results. A report, normally containing only a small fraction of the model outputs – some statistical measures – is provided to users of the ECM. In the evaluation of reporting risk, we assume that the output reflects the company's risk situation well; we focus on the selection and presentation of key figures. As these can have significant influence, they must be assessed in the light of the intended use and the users. This is by far the most difficult part of ECM validation, because the validation team members themselves are influenced by the report content and format. We recommend that this assessment is done by the most senior validation team members.

Finally, we observe that larger ECMs typically have several sub-models; the guidance of assessing model risk sub-categories applies to each of these. We discuss specific validation issues of typical ECM sub-models for a property / casualty insurance company. It is noteworthy that a positive assessment of each sub-model is insufficient; their aggregation needs to be assessed as well.

Section I: Introduction

Economic Capital Models (ECMs) have become a ubiquitous tool for larger, complex insurers. These models are used in three main ways: to inform the process for managing risks and optimizing returns on risk, to allow management to determine an appropriate level of capital to hold for their retained risks, and to satisfy regulatory requirements. These three uses overlap to greater and lesser extents in different territories. But for all three uses, it is extremely important that the underlying concepts are consistent with the intended application of model results, and that the model produces results which are consistent over time. The process of assuring both types of consistency is called model validation.

Model validation is required for many overlapping reasons. If regulators or senior managers base decisions on the output of an ECM, then they should understand the model's assumptions, restrictions and output – and should ensure that the model is suitable for the specific application. But since there are few commonly accepted standards for model validation, the validating party often receives an overwhelming amount of documentation – without the benefit of a standard format or framework for this material. The complexity of modern ECMs demands a more structured validation process that can be understood by decision makers.

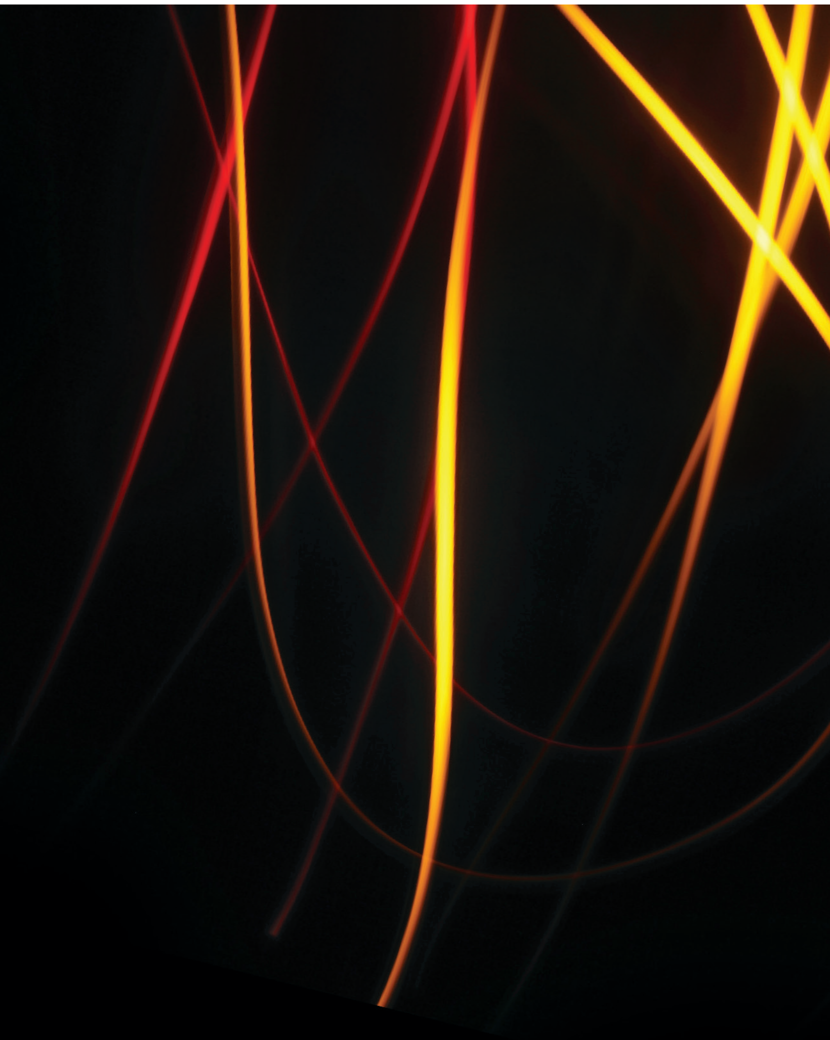
Other fields provide a wealth of experience with validation processes from which we can learn. In software development [1], a lot of effort goes into test procedures; in engineering, approval of a new jet engine requires a stringent validation process; and in the pharmaceutical industry, getting a new drug approved requires an extensive acceptance process. These disciplines have recognized the need to standardize the process of validation and approval. Such standardization also aids the process of communicating the results of validation to the executives who are the ultimate audience.

After initial approval, jet engines and pharmaceuticals are subjected to periodic quality control tests to assure that initial specifications are being maintained. However, here economic capital models deviate drastically from engines or drugs. Although an ECM can be validated before being brought into use, it will be changed, modified and adapted with almost every subsequent use. This means that post-deployment quality control must be almost as robust as the initial validation.

Our hope is that a more standardized validation process will have various benefits: it will create a more objective and less people-dependent result, allow more stakeholders to understand the model's capabilities and restrictions, improve the efficiency of the validation process, and lead to more concise documentation.

WHY VALIDATE?

As statistician George Box famously said: "All models are wrong… but some are useful." The purpose of validation is to assess the degree of model risk, and ensure that this is not so severe as to prevent the model being useful.



Section II: Motivations for model validation

Many companies see the development and use of an ECM as good business practice, the pinnacle of an Enterprise Risk Management (ERM) program to identify, manage and mitigate their risks. Validation of that model is vital for reasons of internal credibility as much as external oversight.

But external stakeholders are also important. In their approach to ECM validation, insurers and reinsurers must consider the views of regulators and rating agencies, and be mindful of the experience other industries – such as banking – have had validation processes. We review a few examples to demonstrate the importance of the subject and to make it clear that more thought is required if the industry is to obtain maximum benefit from ECM validation without incurring crippling costs.

Lessons from the banking industry

Banks pool funds provided by households and firms and invest them either directly (i.e., lending) or through securities markets. Two features distinguish banks from insurance companies: banks' liabilities are typically payable on demand, which creates liquidity risk; and loans comprise a major part of bank assets, which exposes banks to credit risk.

Under the current Basel II regulatory regime banks have the option, subject to certain conditions, to use their own parameter estimates in the calculation of the required capital for credit risk. Although full-fledged ECMs for credit risk are not recognized by bank regulators in the first pillar of Basel II (capital adequacy), intensive discussion is currently ongoing as regards requirements for the inclusion of liquidity risk in the new capital adequacy regime (Basel III).

In order to provide more detailed guidance on validation to national and international supervisors, the Basel Committee formed the **Validation Sub-Group of the Accord Implementation Group**. The Sub-Group defines validation as follows:

> *"In the context of rating systems, the term "validation" encompasses a range of processes and activities that contribute to an assessment of whether ratings adequately differentiate risk, and whether estimates of risk components (such as probability of default, loss given default, or exposure at default) appropriately characterize the relevant aspects of risk." [2]*

This definition and the validation principles proposed by the Sub-Group contain three key elements.

First, the objective of validation is *"assessing the predictive ability of a bank's risk estimates and the use of ratings in credit processes"* (Principle 1 in [2]). This may look obvious, but there is a risk of forgetting it when faced with the complexity of a bank's systems and processes, complexity that does not allow the straightforward application of statistical tests of "good fit."

Second, there is no single "best" validation method, and banks need to use a set of techniques. In the context of credit risk, for example, back-testing may prove inadequate for portfolios where historically defaults have been few. More generally, validation must include both quantitative and qualitative elements.

Third, statistical approaches must be complemented by an assessment of model "governance" (controls, documentation, use test, and so forth). A particularly important aspect of model governance is the process for reassessing the various risk parameters when actual outcomes materially diverge from internal estimates.

Insurance regulation

Solvency II

The legal requirement for model validation in Solvency II is formulated in Article 124 of the Solvency II Directive of November 2009:

> *"Insurance and reinsurance undertakings shall have a regular cycle of model validation which includes monitoring the performance of the internal model, reviewing the on-going appropriateness of its specification, and testing its results against experience. The model validation process shall include an effective statistical process for validating the internal model which enables the insurance and reinsurance undertakings to demonstrate to their supervisory authorities that the resulting capital requirements are appropriate. The statistical methods applied shall test the appropriateness of the probability distribution forecast compared not only to loss experience but also to all material new data and information relating thereto. The model validation process shall include an analysis of the stability of the internal model and in particular the testing of the sensitivity of the results of the internal model to changes in key underlying assumptions. It shall also include an assessment of the accuracy, completeness and appropriateness of the data used by the internal model."*

This legal requirement prompted the regulator CEIOPS (precursor to the pan-European regulation EIOPA) to issue further guidance in the October 2009. Pages 143-169 of CEIOPS' **Advice for Level 2 Implementing Measures of Solvency II: Articles 120 to 126, Tests and Standards for Internal Model Approval (former Consultation Paper 56)** include guidance on validation. In this article, validation is defined as *"a set of tools and processes used by the undertaking to gain confidence over the results, design, workings and other processes within the internal model"*.

The document makes clear that CEIOPS *"considers that, because of the broad scope of the internal model, the validation does not only apply to the calculation kernel to calculate the SCR, but shall encompass the qualitative and quantitative processes of the model. Examples of the areas of the internal model that need to be validated shall include at least:*

a. Data

b. Methods

c. Assumptions

d. Expert judgment

e. Documentation

f. Systems/IT

g. Model governance

h. Use test"

CEIOPS also notes that "*this is not an exhaustive list.*"

Unfortunately, the categories listed are overlapping — e.g., assumptions can relate to methods and data, and documentation can refer to any (and ideally all) of the other topics listed. Furthermore, the comment that the list is not exhaustive makes it difficult for insurance companies to derive clear guidance for their ECM validation process.

To assess market experience we can follow some of the comments of the FSA in the United Kingdom. It has the highest number of internal model pre-approval requests of any European regulator and, due to its experience of the ICAS process since 2005, has perhaps the broadest and deepest experience of internal model vetting of all European regulators. In May 2012 Julian Adams, Director of Insurance at the FSA, stated that insurers were struggling with model validation and as a result that work in this area was falling behind other areas of model development. *"In some cases, [the] scope [of the validation work stream] is too narrow while in others [the] work is simply incomplete. This has led to firms not being properly able to identify critical issues, which has in turn sometimes impeded our review work."*

> NEED FOR GUIDANCE
>
> "In some cases, [the validation] scope is too narrow while in others work is simply incomplete." "…some of the validation policies we have seen have been so vague that we have not been able to draw any assurance from them."
>
> −Julian Adams,
> FSA Director of Insurance,
> May 2012

Later in May 2012, Adams elaborated: *"The first and overarching point is that some of the validation policies we have seen have been so vague that we have not been able to draw any assurance from them. More specifically, we have seen a number of validation policies where judgments as to the materiality of certain elements have been made without necessarily being justified, or with little or no supporting analysis. We would expect decisions about materiality thresholds for validation to be more clearly articulated and justified. To clarify, the level of detail should reflect the materiality of the elements of the model; we are not asking for a detailed justification where it is clearly evident that the element of the model is immaterial."* The key element is that validation needs to be tied to business use, Adams went on: *"The importance of a challenging governance process has also been highlighted in this area. During our review, we saw a validation policy which had been through the full governance process within the firm, and yet when we reviewed it a significant number of areas had been missed, including, for example, ad-hoc triggers for additional validation, the appropriateness of validation tools, the validation of partial internal model and standard formula integration techniques, and independence between the design and validation teams. The governance process needs to add value and challenge, and not simply be a box-ticking exercise through committees."*

The intention is clear and simple, but the practice is opaque and difficult. Prescriptive validation rules do not guarantee an efficient model, and indeed may encourage focus on box-ticking rather than fitness for purpose.

IAIS guidance

While our work on validation is in large part motivated by the Solvency II developments in Europe, the topic is a live issue throughout the world. In the global insurance industry Solvency II is but one manifestation of a broader regulatory movement towards principle-based regulation, in which standard risk-based capital formulae may be replaced by "approved" internal capital models. The International Association of Insurance Supervisors' (IAIS) guidance paper ICP17 paragraph 17.6.9 states:

"By its very nature a standardised approach may not be able to fully and appropriately reflect the risk profile of each individual insurer. Therefore, where appropriate, a supervisor should allow the use of more tailored approaches subject to approval. In particular, where an insurer has an internal model (or partial internal model) that appropriately reflects its risks and is integrated into its Risk management and reporting, the supervisor should allow the use of such a model to determine more tailored regulatory Capital requirements, where appropriate. The use of the internal model for this purpose would be subject to prior approval by the supervisor based on a transparent set of criteria and would need to be evaluated at regular intervals. In particular, the supervisor would need to be satisfied that the insurer's internal model is, and remains, appropriately calibrated relative to the target criteria established by the supervisor (see Guidance 17.12.1 to 17.12.18)."

Despite the caveat *"It is noted that the capacity for a supervisor to allow the use of Internal models will need to take account of the sufficiency of resources available to the supervisor"* the clear guidance is that internal capital models are to be encouraged, but they must be adequately validated.

The IAIS recommends a minimum of three tests for model approval: statistical quality test, calibration test and use test. As seen previously, in a practical implementation consistent with these principles, Solvency II, a few more tests were added. The IAIS also encourages internal and external model validation in paragraph 17.13.6 of ICP 17:

"In addition, the insurer should review its own internal model and validate it so as to satisfy itself of the appropriateness of the model for use as part of its risk and Capital management processes.[35] As well as internal review, the insurer may wish to consider a regular independent, external review of its internal model by appropriate specialists."

Again there is a caveat: *"Validation should be carried out by a different department or personnel to those that created the internal model to facilitate independence."*

Clearly, model validation is core to the approval process — but how can this be achieved in a rigorous way without imposing an impossible burden on insurers and reinsurers? The problem was illustrated recently when a leading insurer, Hiscox, publicly withdrew from the internal model approval process for Solvency II, citing overly complex model validation rules and heavy-handed implementation of those rules by the local regulator. Earlier, Hiscox's Chairman and CEO had openly expressed frustration at the 5000 pages of documentation required, a volume of paper that defies meaningful system validation and audit.

Other regulatory regimes

Even when regulators choose not to implement an internal model regime, there is an implicit drive to use an internal model to estimate capital under Own Risk and Solvency Assessment (ORSA) standards. In Solvency II, the ORSA requires insurers to confirm that the standard formula is appropriate for their capital needs and that they understand their drivers of capital and their sensitivities. While in the U.S., the NAIC version of the ORSA does not require reference to the Risk Based Capital (RBC) requirements, the ORSA Guidance Manual specifically requests that the insurer state their process for model validation. In the United Kingdom pre-Solvency II, the **Individual Capital Adequacy Standards (ICAS)** regime, in some aspects a proto-ORSA, made no demand for internal capital modeling; but the majority of companies of any size saw that an internal ECM was the only logical response to the ICAS demand that companies make an assessment of their own capital requirement. ICAS and ORSA do not prescribe a rigorous validation regime, but companies need to demonstrate that an appropriate validation of the model has been performed.

Rating agencies

Rating agencies too are increasingly expecting larger, more sophisticated (re)insurers to demonstrate use of an internal capital model. Questions about economic capital models and (ERM) now appear on A.M. Best's Supplemental Rating Questionnaire. Standard and Poor's (S&P) even allow companies with a "strong" ERM rating to submit their models for approval; they offer the carrot of using modeled capital results in the rating calculation, the weight given to such results depending upon the assessed credibility of the model.

A.M. Best

This agency's approach does not specifically use the word validation, but A.M. Best does describe an approach to ECM assessment that applies many of the elements of validation, which they call the "key elements of Internal Capital Model process and results":

- Analytical framework and granularity, flexibility, computability, tractability and auditability of the model
- Assumptions and scenario testing
- Timeliness and availability of data
- Applicability and relevance of data
- Sample length and relevance
- Time horizons
- Risk metrics used (VaR, TVaR, CTE, etc.)
- Correlations and dependencies
- Operational, Strategic and Emerging risks
- ICM and management decision making
- Disclosure (internal and external)
- Parameter error / model error and model implementation error considerations
- Internal and external "audit" findings

Standard and Poor's

S&P first introduced an ERM review as part of its rating process in 2005. In January 2011, its criteria were updated in line with regulatory developments to include review of ECMs. The Level III process — only open to those insurers which have a minimum S&P ERM rating of "strong," pass the "use test" with strong strategic risk management, and are willing and able to share sufficient information with S&P to allow assessment of the ECM — can adjust the overall assigned rating level upwards or downwards based on S&P's view of the ECM.

S&P assigns an "M-factor" which reflects its level of confidence in the value calculated for required capital by the company's ECM as compared to the agency's own risk-based capital calculation (RBC). The highest M-factors are allocated to ECMs that S&P views most favorably, and a zero rating will be given to those models not viewed as credible. S&P calculates the M-factor by reviewing a set of financial and non-financial risks such as credit, market, insurance and operational risk. Each is assigned a score of "basic," "good," or "superior" for each of five criteria:

- Methodology
- Data quality
- Assumptions and parameterization
- Process and execution
- Testing and validation

Testing and validation covers the insurer's "*methodologies, interpretation of data and ECM outputs into its ERM program*." S&P states that it views this as one of the most important areas of capital modeling, but limited guidance is given on how the agency assigns scores. Importantly, an ECM will not be deemed credible if insufficient validation has been carried out.

S&P expects insurers' testing of their ECM to cover parameters, assumptions and dependency structures, including stress and scenario testing at "*appropriate confidence levels*." The company has discretion as to choice of scenarios, but adequate justification must be provided. S&P offers several examples of stress tests, all of them historical scenarios:

- Market and credit risk: the 1987 and 1929 crashes as well as 2008
- Mortality and morbidity: the influenza pandemic of 1918 – 1919
- Natural catastrophe: 1813 New Madrid earthquake, 1923 Tokyo quake, 1938 Great New England Hurricane

Overall, the message is that S&P has aligned its criteria with ongoing regulatory changes, including Solvency II, to examine a company's development and use of ECMs. The company's approach to testing and validation will be key to achieving a positive outcome.

Section III: Structuring a validation process: what, who, how

ECM development often takes many man-years; proper validation may require a significant amount of time and involve a number of professionals. Consequently, it is important that the validation process be clearly structured to ensure application of a common standard, with no gaps in the process. To design an appropriate process we must be clear about what the goal is, identify who will own the process, and specify how it will be conducted and documented.

What: assessing the level of model risk

A logical process must be designed with its ultimate goal in mind; in the case of model validation, the goal is to assess the level of model risk. As a side effect, validation results often lead to suggestions for improvements — and consequently to a reduction of model risk.

But what exactly do we mean by model risk? In broad terms, model risk comprises various forms of errors in, and/or inadequate use of, the model. More explicitly, we define the following five sub-categories:

Conceptual risk

Risk that the modeling concepts are not suitable for the intended application.

(We use the terms "appropriate / inappropriate" to denote instances that are "suitable / not suitable for the intended application.")

Implementation risk

Arises from two sources:

Risk that modeling concepts are not implemented appropriately, i.e. the *wrong algorithms were chosen to implement specified concepts*

Risk that the implementation contains errors, i.e. appropriate algorithms were chosen, but the *software contains coding errors or "bugs"*

Input risk

Risk that input parameters are inappropriate, incomplete or inaccurate

Output risk

Risk that key figures and statistics produced by the model are *too sensitive with respect to input parameters or do not support the business purpose*

Reporting risk

Risk that the representation of the output for the business users is *incomplete or misleading*

We observe that because reports are driven by the intended use, reporting risk is related to what is called the "use test" under Solvency II. In a use test the firm has to prove to the regulator that reports produced from the model are used in a business process. Some people consider this to be outside the scope of a model validation, but we consider it to be a crucial aspect of validation.

It is very important to have a clear and practical definition of model risk as this provides the basis for the communication among those involved in validation. Some authors [3] do not offer a conceptually clear separation of the different model risk sources: usually, these authors do not define implementation risk clearly and do not separate between output risk and reporting risk — leaving the reader puzzled about what really needs to be done in a validation. Others, like the Federal Reserve Bank's paper on model risk management [5], are conceptually clean, but avoid giving explicit guidance. An exception is Lloyd's of London's guidance to managing agents on model validation in Solvency II [4]: although it lacks a definition of model risk, and fails to address implementation risk, this is one of very few papers that provide explicit guidance.

Who should validate?

We believe that internal audit is the natural owner of the validation process. This does not mean that internal audit cannot draw on experts from risk management functions, but it will have to establish a validation policy and ensure that it is followed.

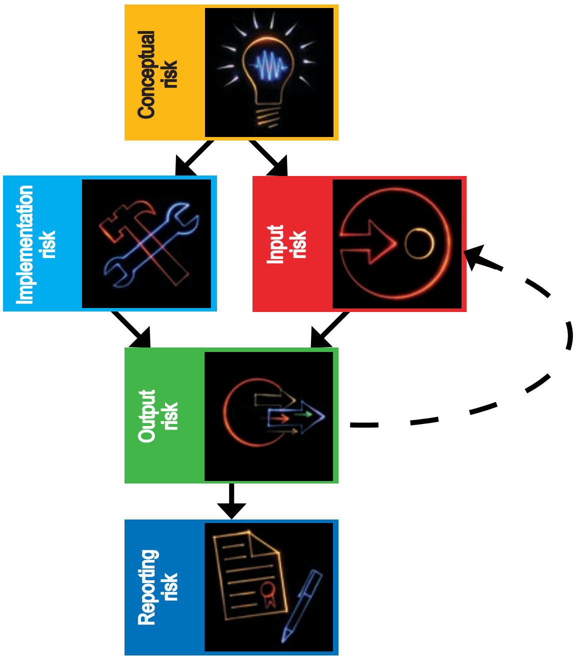
Some confusion about the owner of the validation process can be seen in the literature. For example, the precursor of the pan-European regulator EIOPA, CEIOPS [7], developed a short list of what a validation policy should contain, which mentions (under the heading of governance) that the risk management function should be responsible for model validation. However, in the same paper CEIOPS stresses that independence within the model validation process is essential. We consider the latter correct. But in our view this implies that the risk management function cannot perform the validation, assuming that this function developed the model.

How: an overview

The validation process should follow the major subcategories of model risk mentioned above. At a high level, we note a few process considerations.

Dependencies among model risk subcategories

To establish an efficient process it is important to notice that there are dependencies among the subcategories of model risk, as shown in the diagram at right.



The consequence of these dependences is that if the validation for a risk subcategory fails, then there is no need to validate the dependent subcategories.

Sub-models

If the model is split into sub-models, then the logical flow above applies for each sub-model.

It is important to notice that if the model is broken up into sub-models, the aggregation must be considered as a sub-model which also has to be validated according to the same standards. Even if each sub-model makes perfect sense in isolation, this does not imply that their aggregation has been done (or even can be done) consistently.

Vendor models

The validation of vendor models should follow the same lines and standards as internally developed models. Using a model or sub-model built by an external vendor can save development and implementation time, but it cannot compensate for lack of risk management knowledge.

We therefore disagree with Aon's push to have regulators accept standardized vendor documentation as a form of validation (see Natural Catastrophe risk below and [6]). The motivation for Aon's proposal was to streamline the validation process. We agree that efficiency in the model validation process is important, but it should not come from validation short cuts; rather, it should follow from applying more standardized procedures.

The validation report

The deliverable of the validation process is the validation report – not to be confused with model documentation (it is part of the validation process to check the model documentation). Rather, the validation report lays out the degree to which each sub-model has been checked, and the results of the assessment.

We recommend the following categories for classifying the depth of validation performed on a given sub-model:

- Superficial, further validation required
- Adequate, no further validation effort required
- Adequate, but ongoing validation required

This will enable identification of areas requiring more careful inspection in subsequent validation cycles.

The report should also indicate precisely what checks were made on each sub-model, for each type of model risk. In the following section of this paper we will propose a framework for these checks.

With this information in hand, the results of the validation for each sub-model should be classified as one of two states:

- Inadequate, requiring change or improvement
- Accepted

Again, given the complexities of economic capital modeling, there is no simple way to aggregate individual sub-model assessments to yield a single score for the model; instead, the aggregation itself must be considered following the categories of model risk listed above.

A report of this nature is rich, yet simple enough to be easily interpreted by management.

Re-validation

Finally, we note that the validation process is necessarily an ongoing effort. It is likely, due to resource constraints, that an initial validation will not cover all model risks equally well. But even if the model is kept stable, there will certainly be new input data from time to time. Thus at a minimum the input risk must be regularly reassessed.

Section IV: Proposed validation process

We structure our proposed validation process along the different categories of model risk. More complicated models are usually composed of several sub-models, each of which must be validated individually – and then, as discussed above, integration of the sub-models must be addressed separately.

1. Conceptual risk

The conceptual risk of an economic capital model cannot be evaluated without knowing its purpose: who will use the model, and for what applications.

Purpose of the model

Often we see that the purpose of the model is treated as an introduction to the model documentation: a brief reminder of why the model was built. We consider this inadequate, as no validator can assess the suitability of the modeling concepts without knowing about the intended application and the users.

Rather than describing in general terms the application setting, we recommend explicitly documenting the intended application of the ECM by including the relevant sections from underwriting, investment or risk management guidelines which make reference to the use of the ECM output. It should be considered a warning signal if it is not clear from these texts which decisions the model outputs support, and to which degree the users can supersede or augment the model outputs with qualitative information or judgment.

Of course, ECM outputs serve as quantitative decision support; management needs to include other sources of information, apply judgment, and possibly challenge model results. By no means do we suggest an automated decision making based on model results, but the role of the ECM in the decision process has to be documented.

1a. Application areas

The validation team should expect to find one or more of the following application areas being mentioned:

Capital management, which includes applications such as

- Maintaining retained risk at a level that is supported by capital
- Assisting with decisions via capital allocation, capital budgeting & return on risk
- Making strategic choices regarding business mix that produce superior risk adjusted returns
- Strategic asset allocation

Risk management, which includes applications such as

- Identification of risk concentrations to improve ability to avoid large losses
- Controlling processes over total risk for business units or product lines or territories
- Understanding, communicating and tracking the risk profile

Performance management, which includes applications such as

- Measurement of risk adjusted returns
- Designing strategies for business units to produce superior risk-adjusted business returns
- Goal setting, management and monitoring of risk-adjusted performance
- Risk-adjusted incentive compensation

Product management, which includes applications such as

- Applying uniform standards across a large organization for product risk and returns
- Product pricing that determines a fair price for the risk taken
- Large account pricing

1b. Model users

At this stage of the validation we need only verify that the reports are addressed to a well-defined audience; it need not be assessed whether these are appropriate users of the information. The latter can only be assessed under the reporting risk / use test section, as we need to assess the presentation of the content to decide whether it is appropriate for these users.

Regulators are concerned with both purpose and users because they want to know whether the model results have a significant influence on the business. The argument is that the company will make more effort to model appropriately if its business depends on it. CEIOPS, for example, states in their advice for level 2 implementing measures on Solvency II [7]: "*Insurance and reinsurance undertakings shall demonstrate that the internal model is widely used in and plays an important role in their system of governance.*" The regulators should not be relying upon the model if the management is not.

This introduces a paradox that we have not seen regulators address. The regulators want companies to use their risk models widely, to ensure that they take care to build good models – thus, hopefully, minimizing the model risk. But if a model is used too widely, then even a modest amount of model risk could be devastating for the company. Even worse, if the regulators prescribe a model (e.g. the standard model for Solvency II) that becomes widely used across the market, model risk can lead to systemic risk.

We believe that the validator should check whether the model is in use (checks 1a and 1b mentioned above), but should not judge the breadth of model application. Indeed, any application outside the originally intended purpose creates model risk. Thus, we disagree with CEIOPS' encouragement to use a model as widely as possible: it is more important to ensure that the model is used *appropriately*.

Concepts and their limitations

With the purpose of the model understood, we turn to conceptual challenges: which risks need to be modeled and which modeling methods should be applied to model them. It is important that the question of which risks are material enough to require modeling be answered first, independently of the modeling method. Often the first question is answered implicitly by the choice of the modeling methods. We consider this to be wrong: the starting point needs to be which risks we *should* model rather than which risks we *can* model.

Choosing the risks that need to be modeled is absolutely central in terms of model risk. Once the choice of which risks to include choice is determined, the model's output will not reflect any risk which was excluded.

We should not assume that we have a quantitative model to answer the materiality question; materiality question has to be answered by experts and can be linked to ORSA. Some authors propose subjective risk ranking or materiality tests [4]. We consider such tests problematic. Applying a single risk statistic (e.g. by coefficient of variation, standard deviation, or tail metrics) to linearly order the risks assumes that all risk factors have the same influence on the model. And if different statistics are used for different risk factors, then it is unclear how they can be compared.

1c. Which risks?

Check the **process used to determine which risks need to be modeled.** Ensure that this makes use of the expert judgment of the business leaders. A warning signal should be raised if the modeled risks are implicitly determined by the risk modeling methods, or by a risk ranking derived from a test statistic.

Once validation step 1c is complete, all subsequent steps of the validation refer only to the modeled risks. Any materiality questions which follow are conditional on the materiality threshold for risk selection. Consequently, model validation approaches which do not clearly separate the conceptual risk of risk selection and other validation steps are bound to be imprecise in answering the materiality question.

1d. Modeling methods

Most models are built largely by assembling existing concepts. There is a value in using existing modeling concepts, as their advantages and limitations are more broadly understood. Validating the selection of modeling methods includes the following checks:

- **External references.** If existing modeling concepts are used, they should be documented by reference to a publicly available source. A warning signal should be raised if a company claims that all of their modeling concepts are proprietary.
- **Documentation of how the modeling pieces are connected and why they can be used together.** Sometimes this is falsely referred to as the aggregation. The aggregation model documentation must contain arguments of sub-model consistency, but this is by far not the only place where this issue arises: any subsequent processing units in a model must ensure that the output of one part of the model can be consistently used by the following one. For example: if a company decides to model claims on an annual aggregate level and for the same model describes a detailed per-risk non-proportional reinsurance model, this is inconsistent.
- **Documentation of the limitations of the concepts.** It cannot be overemphasized how important it is to validate that the limitations are documented. Business users must be informed about the limitations; also, those who implement the model are usually less knowledgeable about model concepts, and may be unaware of the limitations that should be checked in an implementation. Therefore, insufficient documentation of conceptual limitations can lead to significant implementation and reporting risk. (For example, if the methods do not provide tail dependence, this should be noted – otherwise, a user could interpret the "excellent" diversification in the tail as a business effect. Similarly, if an economic scenario generator does not provide claims inflation, or if the model does not use the claims inflation it provides, the user must be made aware of this limitation.)
- **Vendor model concepts.** Use of vendor models forces reliance on the vendor's information about their modeling concepts. In this case, documenting limitations is particularly challenging; vendors have little interest in exposing the limitations of their products, and so the modeling team must derive these themselves or collaborate with a knowledgeable service provider. In any event, the validation should verify whether vendor modeling concepts and limitations are declared as vendor provided or self-derived. A warning signal should be raised if a substantial part is vendor provided. In this case the validation team needs to verify that the modelers truly understand the models. This is a challenging task given that most validation teams are small in comparison with the modeling teams and thus it is not easy for the validation team to match all the skills of the various modeling teams.

> **CONCEPTUAL**
>
> *Checking for conceptual risk must start with an understanding of the ECM's intended application. In this context, the validation team should ensure that items selected for modeling are appropriate; that the methods selected are suitable; and that reports are addressed to a well-defined audience.*

2. Implementation risk

An economic capital model is implemented as software; thus, a realistic assumption is that it contains errors. Implementation risk may well be the most underestimated model risk.

Most actuaries and risk managers want to implement their models in their departments, which are usually not staffed with IT professionals. Their argument is that risk management domain knowledge is more important than software engineering knowledge. We believe that both are equally important. Risk professionals often like systems in which they can easily change the models (e.g. spreadsheets and a large section of broadly accepted actuarial modeling tools), due to their unfamiliarity with modern software engineering tools and lack of collaboration with people who understand these tools. It is well documented that the implementation risk of spreadsheet-based systems is heavily underestimated by the business domain experts who implement them; such systems are very hard to test and validate in a professional way [8].

The major issue is the **application of and compliance with best practice software engineering methods.** In many situations the models to be validated are available for detailed inspection, including both documentation of algorithms used and the computer code used for implementation. Nevertheless, our model validation approach will not involve a direct review of computer code. In many situations an analysis of the software code level is unrealistic because it is too time consuming, and would necessitate a benchmark implementation in the form of an independent test implementation or an extensive independent test suite.

> *IMPLEMENTATION*
>
> *Checking for implementation risk is primarily a matter of compliance with best practice software engineering methods.*

2. Implementation checks

- **Risk modeling experts have been involved in the selection of the algorithms** that implement the modeling concepts.

- Check whether **everything in the model development is versioned**: the model code, the reports, the test cases, the test reports. Versioning just productively released versions is not sufficient as it may well be necessary to unwind a change which did not produce the desired result. A warning signal has to be raised if the versioning is missing or is performed by manually renaming files (error-prone and unreliable).

- **Clear accountability for code changes, bug fixes or improvements.** Any change should be initiated by an authorized person, coded by an assigned programmer and assigned to a tester. All three people or teams must be visible in the issue tracker or the code versioning system.

- **Automated test procedures** should be run at regular time intervals (even better, after every new version is checked into the versioning system). A warning signal should be raised if a risk model is only manually tested, as such tests are too infrequent and usually focus only on the new features or new bug fixes. Checking that the most recent changes did not affect correctly running parts of the model is often neglected.

- **Specification of test cases.** Test cases must be specified by the business domain experts and not the implementation specialists.

- **Test coverage reports.** Ideally, the developers provide these reports to the validation team. In the absence of automated testing, the validation team's only way of assessing the level of testing is to check the existence of test protocols and verify that enough time has been allocated to testing– which can only count as a superficial validation of this aspect of implementation risk.

- **Test content.** The validation team must decide which test cases to check in full detail. Ideally, it should be those that have a major impact on the output. But this is not always easy to figure out; a major error in a minor routine may well have as much impact as a minor error in a major routine.

- **Limitations of the algorithms.** Strictly speaking this is covered by the previous point, test content. But since a very frequent flaw in test procedures is that only positive tests are formulated, we make separate mention. Positive test cases check that if the input parameters are correct, then the output is correct. Negative tests must be formulated to ensure that the software catches situations in which the parameters are outside an admissible range. In these situations the software should warn the user or block the calculation.

- **Automatic data loads** from source systems should be tested by integration tests. Again, strictly speaking this is covered by the general check of test content. But the nature of these tests is very different. During development the developers usually work with mock-up data in order to be independent of the source systems. These tests have to be carried out by IT professionals; the validation team simply verifies that they were carried out and that the failure rate was acceptably low.

- **User acceptance testing** is very different from the other forms of testing. It cannot be automated and is usually only performed on stable versions intended for release. Users may have tested real-world scenarios, theoretical worst cases, or just verified that the average model outcome agrees with their expectation. Ideally, test protocols should document which tests the users performed; but unfortunately user acceptance test protocols are often unavailable. In this case, the only thing that can be verified is that the users have been given access to the test environment and enough time to test the new model version. A warning signal should be raised if the users were not involved before the release of a new model version, or if there is no documentation of user acceptance testing.

- **Backtesting and P&L attribution** allow users to gain confidence in the outputs. As insurance models deal with very low frequency data, backtesting is usually not feasible, at least in a probabilistic sense. What can be done is to switch the model from a probabilistic one to a deterministic one: choosing, for example, old realized asset returns, loss ratios and explicit large losses and verifying that if these are used, then the model produces P&L figures in line with the realized ones. There are two issues with this approach: first, the model must produce P&L figures as intermediate outputs (which many ECMs still do not provide). Secondly, while passing this test verifies that the business mechanics of the model are most likely correct, it does not tell us anything about the probability of extreme losses which could threaten the solvency of a company.

It may be surprising that only a fraction of the suggested checks deal with content; most deal with process. Applying sound software engineering techniques and extensive automated tests cannot guarantee error free software, but it can substantially reduce the implementation risk.

INPUT

Checking for input risk involves examination of how both raw and calibrated inputs are selected and handled; validation of any input calibration processes; and benchmarking of the input parameters used.

3. Input risk

One way to express the principles for input validation is simply: internal and external data must be demonstrably appropriate, accurate and complete [9]. This is easy to understand, yet it is very hard to derive explicit guidance from these principles. The three terms characterize partially overlapping issues which sometimes move in opposite directions: more appropriate does not necessarily imply more accurate, more complete does not necessarily imply more appropriate – it can be the opposite. In fact, "accuracy" is very hard to quantify in this context. For these reasons, we prefer an approach that verifies consistency instead of accuracy.

Inputs can be segmented into two classes: raw inputs are items used as parameters for the model without processing the source data, whereas calibrated inputs comprise parameters that have been derived from source data by means of clustering, grouping, averaging, statistical procedures such as distribution fitting, etc.

3a. Types of inputs

- **Confirm that model parameters are clearly designated as either raw or calibrated inputs.**

- **For raw inputs,** verify that the tool does not allow user edits. A warning signal should be raised if raw data can be edited by model users without leaving a trace as this would mean that calibrated inputs are camouflaged as raw data. In effect this is a form of implementation risk, but we mention it here because we did not have the classification of parameters in the previous section.

- **For calibrated inputs,** verify that the data source is well-defined, the calibration procedure is documented and the persons performing the calibration have the required skills.

Generally, more data leads to better calibration. But in the context of models which apply different conceptual frameworks for different phenomena, it is important to verify that the data used for calibration is consistent with the modeling concept. For example, in insurance models, attritional losses and large losses are typically modeled using different concepts for the size-of-loss probability distribution; and both types of modeled losses differ from scenario testing, which normally consists only of a list of possible, extreme parameter values.

3b. Input calibration process

- **Verify that the calibration process uses the data consistently**: data used to derive each calibrated input must be clearly defined in a meaningful way, without overlaps or circularity. For example, data used to calibrate attritional losses must not be re-used to calibrate large losses and vice versa.

  Depending on the integration of any deterministic scenarios with the probabilistic model output, the same issue may appear between large losses and scenarios. If scenario losses are integrated with the probabilistic model results, then large losses which stem from scenarios should be excluded from the large loss calibration.

- If a calibrated input is not derived solely by applying statistical methods, the validation needs to verify that a **peer review process** is in place.

Some practitioners consider conservative assumptions to be a way to reduce input risk. But with raw inputs, by definition there is no option to apply conservative assumptions; and with calibrated data their application introduces opacity. Conservatively estimated parameters also render useless any comparison with historic parameters for the company or the industry. But both of types of benchmarking are valuable instruments for model validation. **We do not believe that introducing conservatism at the input parameter level offers any benefit.** A conservative approach to risk management may well be worthwhile, but such a management approach is not a matter of tweaking model parameters.

3c. Input parameter benchmarking

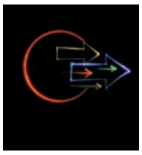
- **Review major changes in source data and input parameter values** since the last validation. Check whether any substantial deviation of input parameters from values used previously has been explained. Ensure that parameter values remaining unchanged from the previous cycle appropriately reflect the underlying data.

- **Benchmark major input parameters industry / peer values.** It will be necessary to make an initial judgment as to which parameters are considered, and then possibly to refine the assessment of these inputs as described in the section on output risk. We propose that if the parameter values that were chosen are in the interquartile range (i.e. the range of 25th to 75th percentile) of the benchmark distribution for this parameter, then the validation only checks whether the calibration procedure is well documented. If the parameter value lies outside the interquartile range of the benchmark distribution, then validation should be more rigorous: in addition to verifying the existence of the calibration documentation, we consider it necessary either to check for a more detailed explanation of the selected parameter values or to verify the actual calibration procedure.

The above guideline puts a lot of emphasis on benchmarking. Currently, there is far too little benchmarking data available. While benchmarking is valuable for validation, still there is research on benchmarking to be done: should overall industry benchmarks be considered – or if not, what is a useful level of segmentation? Which time frame should be used to construct the benchmark distributions? These questions, while interesting, are outside the scope of our current discussion.

OUTPUT

Checking for output risk should validate operational issues, assess the dynamic behavior of the model, and reflect model change analysis.



4. Output risk

We would like to remind the reader that before assessing the output risk, the conceptual risk, implementation risk and input risk should be assessed and deemed acceptable. Validation of output risk is an additional and subsequent process. The assessment of output risk needs to check whether technical experts can interpret the outputs in the context of the intended application. (The presentation and communication of these outputs to business decision makers is dealt with in the reporting risk section.)

4a. Output risk – operational issues

- **Check that the outputs reference the correct input data set and model version.** No one can interpret results without knowing what the outputs refer to.

- **Verify that the outputs can be reproduced.** During an operational run there will be multiple input datasets in use and potentially even multiple model versions. Proper data management is thus imperative: output data needs to reference the input data and the model used. The input data needs to be locked as long as the output data is available to users. A warning signal has to be raised if this is all done manually.

  — For deterministic models, this means verifying that nobody can edit or delete input data sets if they have been used to produce outputs.

  — While this also applies for Monte Carlo models, the output should also reference the random number generator seed(s).

  — If the application runs in a distributed environment, then this computing environment might also have to be referenced in the output; most implementations of distributed Monte Carlo simulations do not guarantee reproducibility across different computing environments.

- **Verify that breaches of input parameter limits** are indicated in the output. There may be good reasons to run a model with an input data set which contains some breaches of parameter limits. These include checking whether the input parameter limits are highly sensitive, or demonstrating that the lack of such limits could lead to outputs which lend themselves to misinterpretations. Such output sets need to be clearly marked.

Once the above operational issues have been checked, then the validation of the dynamic behavior of the model can proceed. This is the most demanding job in the validation process, as it requires both technical expertise and business understanding. The overall goal is to check that the outputs are meaningful.

It is generally accepted that a measurement of the sensitivities of the output key figures with respect to the input parameters yields good insight into the dynamics of a model, but this is only partially true. Even if all derivatives of a functional like the economic capital are zero or very small, there can be significant sensitivities with respect to joint movements of input parameters. But there are usually far too many parameters to measure sensitivities with respect to all input parameters — let alone joint parameter movements. Unfortunately, there is no mathematical procedure to determine which sensitivities should be calculated.

Therefore, business knowledge must be applied to select which sensitivities to study. This introduces output risk: even if the model were a perfect fit to reality and the input data absolutely accurate and complete, it could still be that the outputs of the model are misleading its users: they might not be aware of the dynamics created by some parameter changes. How the dynamic behavior of a model can be explored and verified is still a matter of ongoing research.

In this context, we propose the following checks:

4b. Output risk – dynamic behavior

- Verify that there is **documentation for the selection of input parameters against which output sensitivity is measured.** People with business knowledge should be involved in selecting these parameters, and all involved parties should be aware that a good selection is critical.

- **Check whether output sensitivities are documented.** If the sensitivities are given as estimated values of first derivatives, then the meaning of such derivative estimates must be thoroughly explained. We recommend use of visual as well as numerical outputs.

- Check **materiality of input parameters based on the sensitivities.** The input parameters which are highly sensitive need to be estimated more carefully, documented with more detail and their limits have to be observed more strictly. This is the iterative process referenced in the section on input risk. If input parameters are re-estimated, the output risk must also be re-evaluated. The validation team must apply professional judgment to determine when this input-output dynamic has been addressed satisfactorily.

- Check whether **ranges of output key figures are made available.** Given that most input parameters cannot be estimated exactly, the validation needs to verify that the uncertainty of the output key figures is explicitly communicated. This can be done using confidence intervals around the point estimates.

- Check whether **benchmarking models were used to validate the output** in the sense of comparing with outputs of other, usually simpler, models such as rating agency or regulatory standard models. (It is very likely that regulators and rating agencies will make such comparisons.) The expectation is not that these simpler models yield the same output; otherwise it would not have made sense to build the more complicated internal models. But experts should be able to explain and document reasons for the differences. This is almost like presenting a value proposition for the internal model.

If the model is not an initial version, then output from older model and input data versions might be available for comparison. Often, a new model version is run with an old data set, and then the results are compared with results obtained from the old model version. For a better understanding of the effects of the model changes, the various model changes are applied one after the other. The problem with analysis of change is that it is not independent of the order in which the changes are applied. Hence, if it is used for validation purposes, the documentation must include arguments concerning a meaningful order of changes.

4c. Output risk – model change analysis

- Check that the analysis of change **starts from a validated model and input data set** and documents the **procedure of how the changes are applied** as well as the **rationale** for the selected order of changes.

5. Reporting risk

This is the final step in ECM validation and the most crucial. Yet some people consider this step to be outside the scope of model validation because it is closely related to the use test. We believe that if this step is not included in the validation process, the model's use in the intended real-world application – the most essential point, in our view – cannot be assessed.

While user acceptance tests referenced in the assessment of implementation risk and output risk are done by technical experts, reporting risk concerns the ultimate users of the model output: management. The model itself and the complete output data are usually unavailable at this level. Management committees and meetings use only the reports created from the model outputs.

It must be emphasized that this validation step is not an assessment of company management. The question is not whether management is good enough to deal with the model's outputs appropriately, but instead whether the model's outputs provide useful decision support for management. While regulators and rating agencies must form opinion about management ability, this has nothing to do with the economic capital model the company is using.

At this stage of the validation we can assume that the output figures are correct; now we must understand the users and the intended application to assess whether users can draw meaningful conclusions from the reports at hand. This requires involvement of the most senior members of the validation team; preferably, these team members should have management experience.

The validation team can and should ask the business users whether they consider the reports useful and meaningful decision support, but the validation team cannot simply rely on the users' opinion. User feedback is asymmetrical: a user who is very little affected by the model's outputs is likely to provide weak feedback, but this does not mean that users who are significantly affected by model output provide the most useful feedback. These users often mix the reports' usefulness with the effect on their business. If model results are favorable, then these users are inclined to give a more positive review than they would if results create uncomfortable business pressure.

5. Reporting risk

- Do reports **clearly state to which model and data version the outputs refer**?
- Are business users made aware of situations in which some of the **parameters are outside a comfort range** or even outside the agreed limits?
- Is the **frequency of reports** in line with the decisions they should support?
- Are results are communicated using **institutionally accepted metrics**? Metrics that may capture or describe the risks well, but are not commonly known or used in the company, introduce reporting risk.
- Check whether the report uses any means to **convey how robust the key figures are.** Simply providing point estimates of the key figures does not give enough information to decision makers. They should be made aware of the fact that estimation errors for the parameters, different modeling assumptions, and so forth affect the range of outcomes.
- Do reports communicate the **range of normal business volatility**? Even if the intended application is to decide about the required economic capital based on extreme outcomes, it is important for business users to know the normal business volatility: extreme situations cannot be well understood without knowing what is considered normal. Interquartile ranges can be considered a good measure of normal fluctuation.

> *REPORTING*
>
> *An assessment of reporting risk should review the clarity of reports; ensure that the reports include proper context for the recipients; and confirm that reporting frequency is aligned with the time frame for relevant decisions.*

Considerations for specific types of sub-models

In this section we discuss issues particular to specific classes of risk models that typically form sub-models of an economic capital model. For each type of sub-models, we discuss considerations that affect the assessment of conceptual, implementation, input, output, and reporting risk as described above.

Underwriting sub-model

Underwriting sub-models should capture the uncertainties and dependencies of an insurer's losses – attritional, large individual and catastrophic – as well as premiums and expenses.

The primary **conceptual risk** in underwriting models concerns whether the modeling framework captures the nuances of the lines of business in the insurer portfolio.

- Check whether the model uses different concepts for **long-tailed and short-tailed business** differently. If so, this should be documented.
- How does the model treat **attritional, large individual and catastrophic losses**?
- What is the concept for modeling common **systemic risks** like claims inflation (e.g. medical inflation).

Underwriting model **input risk** centers on the selection of frequency and severity of losses by line of business, dependency strength, projected rate levels, and the parameter uncertainty inherent in these factors.

- Check whether the **selection of parameter values is well documented**. Often they are not purely derived from company data; thus the process of using expert knowledge or benchmarking data to augment the company's data must be documented. This is especially true for parameters of dependency models.
- Trends in loss development and rate change assumptions need to be evaluated in light of company history and also benchmarked against industry movements.

Underwriting model **outputs** can be highly sensitive to trend assumptions for long-tailed lines and correlation assumptions among lines of business.

- Check for **comparisons to prior results**. Sudden movements in mean results or changes in any statistical key figures (e.g. coefficient of variation) should be explained or the validation has to investigate them.

Since underwriting drives the company's potential income, the following is a **reporting risk** issue:

- Check whether the loss potential and loss scenarios are **presented in relation to the underwriting profit**.

Natural catastrophe sub-model

Natural catastrophe risk is typically modeled within a sub-model of the underwriting risk sub-model. For the purposes of this paper we will divide catastrophe models into two classes: "physical-based" and "actuarial-based." The first type are developed using complex physical science, such as seismology and meteorology; what separates them from the second type is the level and nature of information available. Both classes of models provide the end user with loss distribution curves to use within the ECM, but the required skills for validating these two types of catastrophe models are very different. Actuarial-based models can be validated using the "underwriting risk" principles outlined above, but physical-based models present additional considerations.

Physical-based catastrophe models are widely used by (re) insurance companies and are mainly developed by specialized vendors. Vendor-provided models enjoy wide acceptance in the market, in part due to the costs and complexity of building proprietary models. This creates an additional risk for many ECM users; they have to rely on the vendor's model documentation.

As stated above, we do not consider it appropriate to delegate the validation of natural catastrophe models completely to the vendors. First of all, the validation must be put into the context of the intended use of the ECM; but the vendor does not know the intended use and so the vendor's validation can only be a generic one. Secondly, the providers have a commercial interest in their model's validation. Therefore, the insurance company using the model must validate it. This validation will use, but not be limited to, the vendor's documentation.

In general, physical-based catastrophe models consist of five components:

1. **Event sets:** collection of events, either historically observed or stochastically generated
2. **Hazard component:** the science of hurricanes, windstorm, earthquake, flood, fire, etc.
3. **Vulnerability component:** the science of how a given hazard affects the built environment
4. **Insurance exposure data:** policy information
5. **Financial component:** calculation of how losses flow through insurance and reinsurance policies, deductibles, etc.

The modeling **concepts** are well documented by the vendors; consequently, the validation has to

- Assess **whether the internal modeling team is familiar with the modeling concepts**. Usually, this includes checking that the arguments why a particular provider was chosen are well documented and linked to the intended use.
- Model changes can have a significant impact on the results. Thus, it needs to be checked that the modeling team documents the **rationale for deliberately updating the model or deliberately staying with an older version**.
- Checking **whether the model covers all major risks** in the insurance company's exposure: for example, a hurricane model that ignores flood despite significant exposure to such flood is inadequate.

To assess the **implementation risk** we have to

- Verify how rigorous and transparent the vendor is in **communicating software bug fixes and improvements**.
- We expect the internal team to check the influence of bug fixes with their own test cases which are meaningful with respect to their portfolio.

Examples of **input risk** indicators include:

- Hazard component
  - Whether the observed and the modeled events appear to be reasonably overlapping
  - Selection of historical events is appropriate
  - Measures for goodness of fit
  - Choices of data flow interpolation
  - Parameterization of the probability distribution
- Vulnerability component
  - Key drivers to insured loss generation are in line with the portfolio's key loss drivers
  - Have claims data used for developing vulnerability functions been interpreted correctly, e.g. policy conditions?
  - Have damage curve data been fitted appropriately?
- Exposure data
  - Checking whether the risk descriptors which are used by the natural catastrophe model (e.g. building structure type / commercial occupancy type / floor level, cellar, occupancy types) are captured in the source systems or estimated.
- Financial modeling
  - Check whether the flow of loss correctly reflects policy conditions

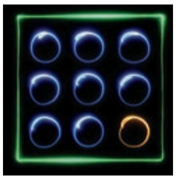
The **output** produced should be tested for sensitivity to change in model's settings to understand the impact these might have on the modeled result. Examples include:

- Was the model run with or without loss amplification, storm surge, etc.?
- Have the modeled results been compared with industry losses, company claims history or multi-model results, and have the differences been explained?

Integration of catastrophe loss distributions into the economic capital model makes it possible to aggregate catastrophe and non-catastrophe exposures. This integration raises the following validation questions:

- Is catastrophe model output directly used in the economic capital model, or is it adjusted? If the output is adjusted, then these adjustments have to be documented.
- Check that the ECM reproduces the catastrophe model output if the non-catastrophe exposures are set to zero.

Reserve risk sub-model



Setting reserves and paying them out over time is a core process of an insurance company. In terms of **conceptual risk**, it is quite common to apply relatively simple methods and then modify the results using expert judgment. This poses some difficulties in assessing conceptual risk:

- Checking which method is applied to calibrate the data is very important. The documentation must include where and who can change the values according to expert judgment. Often outside consultants are involved. In this case, the validation team should verify that the internal modeling team does in fact assess choices made by the consultants and take responsibility for the result.
- Check the method for creating reserve variability. Does this method work with gross, or with net and ceded? Verify that the granularity fits the data.
- Check whether the model deals with correlations.
- Underwriting cycles can have a significant effect on reserves. Hence the validation should note whether this factor is considered.

There are usually several IT systems involved in the reserving process, and the reserve risk model uses their data and results. Hence, the following validation issues arise concerning the **input risk**:

- Has the risk modeling team documented which data sources they are using for calibration?
- Have data sources been merged?
- Is the segmentation which has been applied reasonable and stable over time?
- If aggregations have been applied before using the data, then these have to be carefully documented.
- Are gross, net, and ceded amounts consistently treated, taking into account changes in reinsurance treaty terms?
- Are the adjustments which have to be applied before the data is used for calibrating the model (e.g. claims inflation) documented?
- Trends in claims and reserve data are very important. Consequently, the validation team has to ascertain that changes in key figures (rates of settlement, caseloads, payout lags, etc.) are monitored by the risk modeling team.

For **output risk**, consider:

- If enough data is available for model back-testing and performance testing (e.g. as discussed in [10]) then the validation has to assess these results

Counterparty risk sub-model

Insurers' largest exposures to counterparty risk usually come from their investment portfolio and from their reinsurance receivables.

Conceptual issues include:

- Does the model deal adequately with the difference between the large number of counterparties in the investment portfolio and the small number of reinsurance counterparties?
- Verify that the model includes exposure to reinsurer default after the year the claim is reported
- Verify that the model reflects the correlation between reinsurer default and years with large claims.
- Verify that the model explains well at which level of aggregation counterparty risk is modeled in the investment portfolio, e.g. individual counterparties or even individual instruments of the same counterparty, or portfolios with the same rating class.
- Verify whether the effects of market value changes are included.
- Confirm that the variations of credit spreads are not being double counted by inclusion in the interest rate models in addition to the credit risk models.

Input risk issues:

- Verify that the granularity of the data fits the model. In particular, this is an issue in the investment portfolio.

Investment risk sub-model

Conceptual issues include:

- Verify that the valuation principles are well documented and fit well with the economic scenario generator (ESG) that is used. The valuation principles and the ESG should fit the intended use and time frame.
- If one of the many well-known asset models is not being used, confirm that the rationale for selecting another model is well documented.

Insurer investments are usually modeled according to their similarity with much larger classes of similar investments. In some cases, the investment sub-model will assign individual investments to a group that is modeled as if that group were identical to the broad class of investments. In other cases, a single investment (or group of investments) is "replicated" with a combination of representative amounts of the broad classes.

Input risk issues include:

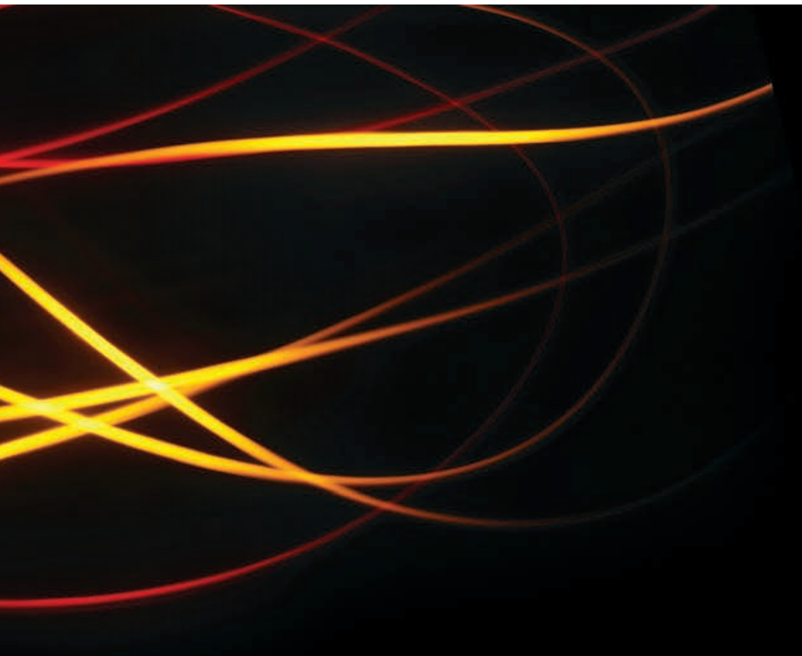
- For the assignment to classes, the degree to which each investment fits naturally into one of the traditional main categories of investments (equities, bonds, real estate, mortgages, loans, etc.) should be documented. For each category, the validation should include review of the "fit" criteria to confirm that investments selected with that criteria would be appropriately represented by the class.
- Compare the actual investments assigned to any class to the class definition. Has the assignment been made on a judgmental basis or were statistical tests performed? For example, if the class to be used was defined as "all investment grade bonds" then the class would have characteristics based upon a mix of bonds with ratings ranging from AAA to BBB. But if in fact the actual investments are almost all AAA, then the actual investments will not have the same characteristics as the model class. Similarly, an actual equity portfolio with a beta of 2 would be skewed compared to a market index class; a portfolio with 25% turnover would be skewed compared to a "buy and hold" class.

Asset and liability matching sub-model

We deliberately put the Asset and Liability Matching (ALM) sub-model in a separate section and not as a subsection of the investment risk sub-model.

- The major issue for the validation is to check whether the liability model and the investment risk model produce consistent outputs. Time aspects, as well as level of detail, should be checked carefully. For example, a very detailed investment model paired in an ALM process with an aggregate liability model should trigger a warning signal.

Since ALM decisions are usually taken by a committee using reports, other validation points are largely those mentioned above under **reporting risk**.



Conclusion

The article proposes five model risk categories which can be embedded naturally in a validation process. The same process must be applied to all sub-models; including the aggregation model. In addition, a standard way of reporting the model validation results is proposed.

The theme of the article is that a more standardized validation process, including the way validation results are being reported, introduces efficiency and objectivity. By treating reporting risks differently from output risks we acknowledge that a validation will never be fully objective.

Appendix: Sample validation checklist

1. Conceptual risk

1a. Application areas

- Capital management
- Risk management
- Performance management
- Product management

1b. Model users

- Verify that reports are addressed to a well-defined audience

1c. Which risks?

- Selection of risks to be modeled makes use of business leaders' expert judgment

1d. Modeling Methods

- External references
- Documentation of how the modeling pieces are connected and why they can be used together
- Documentation of the limitations of the concepts
- Vendor model concepts

2. Implementation risk

2. Implementation checks

- Risk modeling experts have been involved in the selection of the algorithms
- Everything in the model development is versioned
- Clear accountability for code changes, bug fixes or improvements
- Automated test procedures
- Specification of test cases
- Test coverage reports
- Test content
- Limitations of the algorithms
- Automatic data loads
- User acceptance testing
- Backtesting and P&L attribution

3. Input risk

3a. Types of inputs

- Confirm that model parameters are clearly designated as either raw or calibrated inputs
- Raw inputs: verify that the tool does not allow user edits
- Calibrated inputs: verify well-defined data source, documented calibration procedure performed by people with the required skills

3b. Input calibration process

- Verify that the calibration process uses the data consistently
- Verify that a peer review process is in place for calibrated inputs

3c. Input parameter benchmarking

- Review major changes in source data and input parameter values since the last validation
- Benchmark major input parameters against industry / peer values

4. Output risk

4a. Output risk – operational issues

- Check that the outputs reference the correct input data set and model version
- Verify that the outputs can be reproduced
- Verify that breaches of input parameter limits are indicated in the output

4b. Output risk – dynamic behavior

- Documentation for the selection of input parameters against which output sensitivity is measured
- Check whether output sensitivities are documented
- Check materiality of input parameters based on the sensitivities, and if necessary recalibrate these input parameters, iterating until the validation team is satisfied that validation can continue
- Check whether ranges of output key figures are made available
- Check whether benchmarking models were used to validate the output

4c. Output risk – model change analysis

- Check that the analysis of change starts from a validated model and input data set and documents the procedure of how the changes are applied as well as the rationale for the selected order of changes

5. Reporting risk

5. Reporting risk

- Do reports clearly state to which model and data version the outputs refer?
- Business users made aware when parameters are outside a comfort range
- Is the frequency of reports in line with the decisions they should support?
- Are results are communicated using institutionally accepted metrics?
- Check whether the report uses any means to convey how robust the key figures are
- Do reports communicate the range of normal business volatility?

Considerations for specific types of sub-models

Underwriting sub-model

Conceptual risk: does modeling framework capture nuances of the lines of business?

- Long-tailed and short-tailed business
- Attritional, large individual and catastrophic losses
- Systemic risks

Input risk: selection of frequency and severity of losses by line of business, dependency strength, projected rate levels, and the parameter uncertainty inherent in these factors.

- Selection of parameter values is well documented
- Trends in loss development and rate change assumptions need to be evaluated in light of company history and also benchmarked against industry movements

Output risk

- Check for comparisons to prior results

Reporting risk

- Check whether the loss potential and loss scenarios are presented in relation to the underwriting profit

Natural catastrophe sub-model

Conceptual risk

- Assess whether the internal modeling team is familiar with the modeling concepts
- Documentation of rationale for deliberately updating the model or deliberately staying with an older version
- Checking whether the model covers all major risks in the insurance company's exposure

Implementation risk

- Verify how rigorous and transparent the vendor is in communicating software bug fixes and improvements
- We expect the internal team to check the influence of bug fixes with their own test cases which are meaningful with respect to their portfolio

Input risk

- Hazard component
  - Whether the observed and the modeled events appear to be reasonably overlapping
  - Selection of historical events is appropriate
  - Measures for goodness of fit
  - Choices of data flow interpolation
  - Parameterization of the probability distribution
- Vulnerability component
  - Key drivers to insured loss generation are in line with the portfolio's key loss drivers
  - Have claims data used for developing vulnerability functions been interpreted correctly, e.g. policy conditions?
  - Have damage curve data been fitted appropriately?
- Exposure data
  - Are risk descriptors (e.g. construction/occupancy) captured in source systems or estimated?
- Financial modeling
  - Check whether the flow of loss correctly reflects policy conditions

Output risk: test for sensitivity to model settings, such as

- Was the model run with or without loss amplification, storm surge, etc.?
- Have the modeled results been compared with industry losses, company claims history or multi-model results, and have the differences been explained?

Integration risk

- Is catastrophe model output directly used in the economic capital model, or is it adjusted?
- ECM should reproduce cat model output if the non-catastrophe exposures are set to zero

Reserve risk sub-model

Conceptual risk

- Method applied to calibrate data
- Method for creating reserve variability
- Check whether the model deals with correlations.
- Underwriting cycle effects

Input risk

- Has the risk modeling team documented which data sources they are using for calibration?
- Have data sources been merged?
- Is the segmentation which has been applied reasonable and stable over time?
- Documentation for any aggregations applied before using the data
- Are gross, net, and ceded amounts consistently treated, taking into account changes in reinsurance treaty terms?
- Documentation of adjustments applied to data before calibrating the model (e.g. claims inflation)
- Ascertain that changes in key figures (rates of settlement, caseloads, payout lags, etc.) are monitored by the risk modeling team

Counterparty risk sub-model

Output risk

- Assess model back-testing and performance testing (e.g. as discussed in [10])

Conceptual risk

- Does the model deal adequately with the difference between the large number of counterparties in the investment portfolio and the small number of reinsurance counterparties?
- Verify that the model includes exposure to reinsurer default after the year the claim is reported
- Verify that the model reflects the correlation between reinsurer default and years with large claims
- Verify that the model explains well at which level of aggregation counterparty risk is modeled in the investment portfolio
- Verify whether the effects of market value changes are included
- Confirm that the variations of credit spreads are not being double counted by inclusion in the interest rate models in addition to the credit risk models

Input risk

- Verify that the granularity of the data fits the model. In particular, this is an issue in the investment portfolio

Investment risk sub-model

Conceptual risk

- Valuation principles are well documented and fit well with the ESG
- Confirm rationale for selecting a non-standard ESG

Input risk

- Assignment of investments to classes
- Degree to which investments assigned to class have class properties

Asset and Liability Matching risk

- Check whether the liability model and the investment risk model produce consistent outputs; time aspects, as well as level of detail, should be checked carefully

References:

[1] C. Kaner, J. Bach, B. Pettichord, Lessons learned in software testing: a context driven approach, Wiley Computer Publishing, 2001

[2] Basel Committee on Banking Supervision, "Update on work of the Accord Implementation Group related to validation under the Basel II Framework", Basel Committee Newsletter, No. 4, January 2005.

[3] Model Validation Principles Applied to Risk and Capital Models in the Insurance Industry, North American CRO Council, 2012

[4] Solvency II - model validation guidance, Lloyd's, June 2012

[5] Supervisory guidance on model risk management, Board of Governors of the Federal Reserve System, OCC 2011-12, April 2011

[6] Aon Benfield Pushes for 'Simplified Internal Model' for Solvency II Approval, Insurance Journal, 11 Feb. 2011, http://www.insurancejournal.com/news/international/2011/02/22/187529.htm

[7] CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Articles 120 to 126, Tests and Standards for Internal Model Approval, CEIOPS-DOC-48/09, 2009

[8] Raymond R. Panko, What we know about spreadsheet errors, Journal of End user computing's special edition on Scaling up End User Development, Volume 10, No 2. Spring 1998, pp. 15-21

[9] CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Article 86f, Standards for Data Quality, CEIOPS-DOC-37/09, 2009

[10] Ying, Lebens and Lowe, Claim Reserving: Performance Testing and the Control Cycle, Variance Volume 3 Issue 2

Willis Re

MANAGING EXTREMES

Global and local reinsurance

Willis Re employs reinsurance experts worldwide. Drawing on this highly professional resource, and backed by all the expertise of the wider Willis Group, we offer you every solution you look for in a top tier reinsurance advisor. One that has comprehensive capabilities, with on-the-ground presence and local understanding.

Whether your operations are global, national or local, Willis Re can help you make better reinsurance decisions – access worldwide markets – negotiate optimum terms – and boost your business performance.

How can we help?

To find out how we can offer you an extra depth of service combined with extra flexibility, simply contact us.

Begin by visiting our website at **www.willisre.com** or calling your local office.

Markus Stricker
Willis Economic Capital Forum
markus.stricker@gmail.com

David Simmons
Willis Re
51 Lime Street
London EC3M 7DQ
Tel: +44 (0) 20 3124 8917
david.c.simmons@willis.com

Dave Ingram
Willis Re
200 Liberty Street
New York, NY 10281
Tel: +1 212 915 8039
dave.ingram@willis.com

Willis Limited
The Willis Building
51 Lime Street
London EC3M 7DQ
Tel: +44 (0)20 3124 6000
Fax: +44 (0)20 3124 8223

Willis Re Inc.
One World Financial Center
200 Liberty Street
3rd Floor
New York, NY 10281
Tel: +1 212 915 7600