Cass Business School
CITY UNIVERSITY LONDON

Catastrophe Model Suitability Analysis: Quantitative Scoring

NAME : XINRONG LI

STUDENT NO. : 050005179

SUPERVISOR : Dr Andreas Tsanakas

The dissertation is submitted as part of the requirements for the award of the MSc in Actuarial Management.

August 2013

ABSTRACT

This paper investigates the problem of translating subjectively qualitative value into quantitative scoring measurement within MSA Grid. MSA Grid is a summary of expert's subjective judgments regarding to catastrophe modeling evaluation across competing catastrophe models. It is an important component within Model Suitability Analysis (MSA) framework, introduced by a leading reinsurance broker firm Guy Carpenter. The purpose of MSA is to develop client's own view of catastrophe risk management therefore MSA Grid performs as a platform for a particular client to determine which catastrophe model is the most suitable one to customize catastrophe risk based on its exposure portfolio. This paper mainly provides two scoring methods across competing catastrophe models and give rise to conclude the suitable model on the basis of certain criteria.

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my supervisor, Doctor Andreas Tsanakas, and my line manager at Guy Carpenter, Guillermo Franco, for their enlightening advice and helpful suggestions on my thesis. I am deeply grateful for both of them to guide me in the completion of this thesis. I am also deeply grateful to all other colleagues at Guy Carpenter, Delioma Oramas, and all the other tutors in Cass Business School, Doctor Zaki Khorasanee, for their direct and indirect help to me. Special thanks should go to my friends, William Wu and Jing Li, who have put considerable time and efforts into their comments on the draft. Finally, I would like to express my sincere gratitude to my beloved parents who have always been there for me and supporting me all these years.

Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Natural catastrophe events, such as hurricanes, earthquakes and floods, can stress the financial position of insurance and reinsurance companies. For example, hurricane Andrew (1992) caused more than $16 billion of loss and left 11 insurers insolvent (AIR Worldwide, 2012) .Such disasters occur relatively rarely worldwide. Therefore, their relative rarity makes it difficult to estimate losses through standard actuarial techniques due to lack of historical loss data. Estimating future catastrophic losses requires then a specialised tool, which gives rise to catastrophe modelling.

Today, there are three leading firms specialising in catastrophe modelling for the insurance industry, which are AIR Worldwide (AIR), Risk Management Solutions (RMS) and EQECAT. All these three firms have undergone a continual process of development of catastrophe modelling and new models have been launched subsequently for new perils and regions in the world, deployed as versions of their respective software platforms, such as AIR V11 [2012] and AIR V9 [2010]. However, the reliability of the outcomes from catastrophe models depends heavily on the correct understanding of the underlying physical mechanisms that control the occurrence and behaviour of natural hazards (RMS, 2012). Therefore, outcomes from different catastrophe models may result in quite different losses even when carrying out the analysis for the same or similar catastrophic events.

As a leading reinsurance broker firm, Guy Carpenter (GC) has utilised over the years those three models to assist their clients, direct insurers, reinsurers, etc. with risk management of catastrophe events. Also recently, GC introduced the Model Suitability Analysis (MSA)^{SM} framework to collaborate with their clients in developing their own view of risk. One main purpose of MSA is to provide a clear synthesis of model suitability for the client's exposure since there is significant uncertainty on the output of competing catastrophe models. To assess such an uncertainty, MSA contains a component dealing with evaluation of catastrophe models in terms of several tests and summarises all the tested results in a color-coded table, called MSA Grid. The grid has been colour coded with red, yellow and green and most of the tests results are given with qualitative value, e.g. "good", "moderate" ,"poor", etc. This means that the assessments of catastrophe models oftentimes involve experts' subjective judgements. In practice, it is a common issue that particular cat modellers do not have a strong expertise to make a judgement with high degree of certainty. Therefore, the main motivation for this study is to investigate methodologies to interpret qualitative values into quantitative measurements in order to decide which catastrophe model is the most suitable one for a particular client's portfolio. Such a topic is common in practice for decision making in every industry. Cooke (1991) has suggested that this issue should be regarded as uncertainty within the expert's subjective judgement. The theoretical methodology referred by Cooke (1991) concentrates on how to achieve a convinced and suitable MSA Grid in my view. However, this issue is out of the scope and concentration of this paper. Thus, at this stage, this paper aims to provide accessible and simple scoring methods to determine the suitable catastrophe model based on a given MSA Grid in terms of deterministic and probabilistic modelling.

This paper is structured as follows. In Section 2, I would provide a brief introduction of catastrophe modeling and further explanations of the uncertainties involved within catastrophe modeling in order to give the reader a deeper understanding of our targeted problem. In Section3, I would present a brief introduction of the MSA framework introduced by Guy Carpenter and also illustrate the issues of using MSA Grid for decision making. In Section 4, I would present two accessible scoring methods in terms of deterministic and probabilistic modeling as applicable tools for catastrophe models 'comparison. Section 5 focuses on a case study regarding to Turkey earthquake together with the application of two scoring methods in different client's exposure portfolio. Some interesting findings can be investigated in this section. Finally, section 6 concentrates on the conclusions of application of two scoring methods in this paper and illustrates further scope to research on the basis of this paper. I have used Excel and MATLAB to perform the modeling and statistical analysis throughout this paper.

2. THE NATURE OF CATASTROPHE MODEL

2.1 What is a Catastrophe model

Before showing the structure of catastrophe model, we can explore the origin of catastrophe model. Figure 1 exhibits the timeline of development of Catastrophe model.

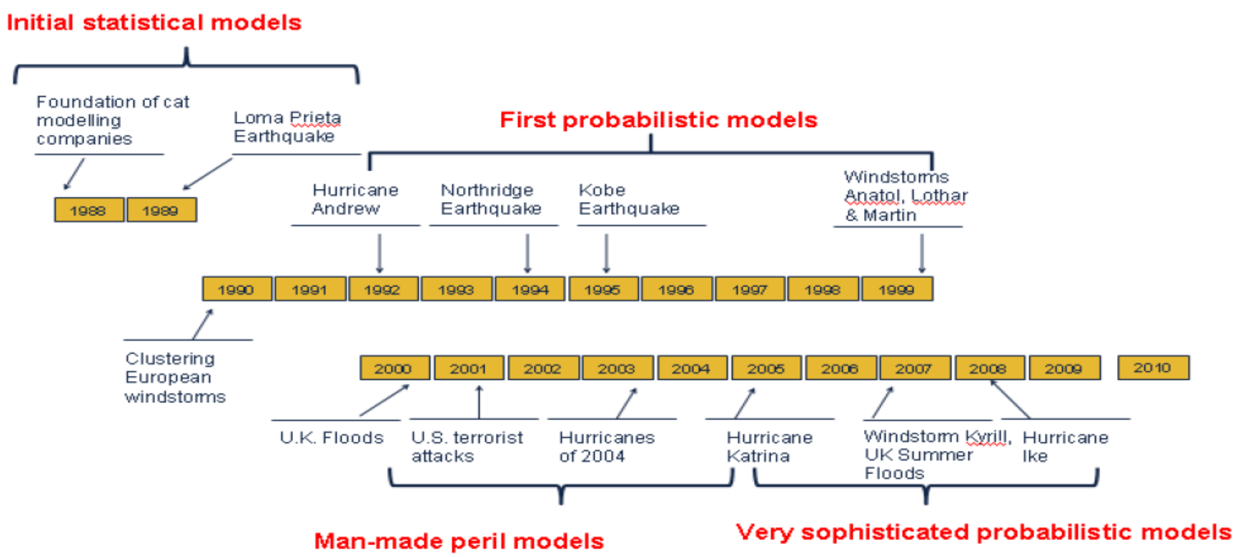

Figure 1: The short story of catastrophe models (Source: Parodi (2012))

Catastrophe models have arisen in the late 1980s accompanied by the foundation of the first catastrophe modelling firm. The techniques used in that period were mainly based on scientific studies of natural hazard measurements and historical occurrences with advances in information and geographic information systems (RMS, 2008). As Hurricane Andrew occurred in 1992 and resulted in unprecedented losses, the first probabilistic models have been driven to become the most appropriate way to manage catastrophic risk (RMS, 2008). However, Hurricane Katrina in 2005 exposed the inadequacies of first generation catastrophe models so that more sophisticated probabilistic models have appeared in the meantime. Every large catastrophe event can force to enhance and refine catastrophe models further so could be possible to say that catastrophe models constantly play a catch-up role with the reality.

Today, catastrophe models are more mature and prevalent throughout the insurance industry, assisting insurers and reinsurers in managing natural perils and man-made catastrophes across the world. Although there are minor variations in the break-down structures across different catastrophe models, the standard catastrophe models for natural hazards can be divided into the following three modules (Parodi, (2012)):

- Hazard Module
- Vulnerability Module (Engineering Module)
- Financial Module

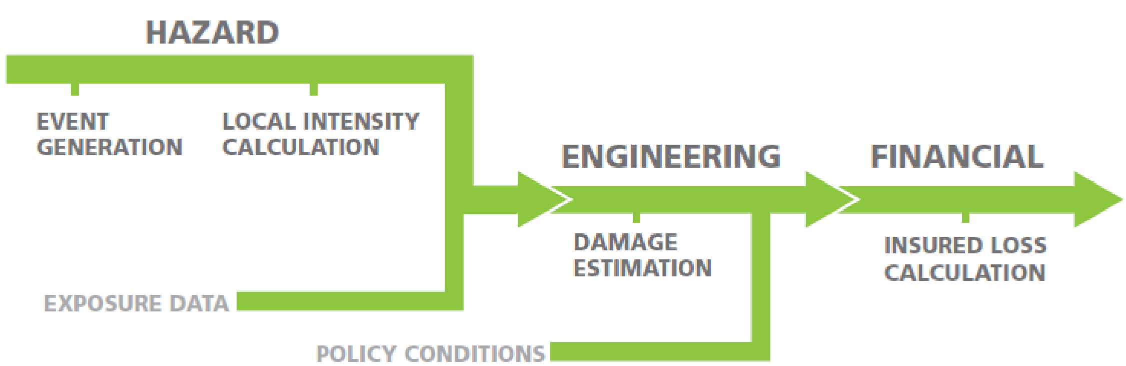

Figure 2 : the structure of catastrophe models (Source: AIR Worldwide (2012))

Figure 2 illustrates the framework of catastrophe modelling together with the inputs from clients (exposure data and policy conditions). Some detailed descriptions regarding to each module are given as follows:

Hazard Module

This module aims to answer the questions: what's the geographic location of future events likely to occur? How large or severe are the events likely to be? And how frequently are they likely to occur? To answer the above questions, we firstly need to produce a large catalogue of potential catastrophe events through computer simulation and secondly calculate the intensity at each location across a geographical area at risk (Parodi, (2012)). The intensity here can be expressed, for example, for hurricanes in terms of wind speed or storm surge height; for earthquakes in terms of degree of ground shaking.

Vulnerability Module

This module focuses on investigating more detailed information, such as the level of building damage expected, on the properties that are exposed to simulated catastrophic events. The level of building damage expected can be estimated as a function of different level of intensity of the event, which is the so-called damage functions generated by region-specific and vary by a property's susceptibility to damage from specific peril, e.g. earthquake ground shaking or hurricane winds (RMS, 2008). For financial analysis, the level of building damage is ultimately measured in terms of a damage ratio, the ratio of the average anticipated loss to the replacement value of the building, ranging from 0% to 100%, or total loss.

In addition, the entire building consists of various components, such as structural components including beams and columns, non-structural components, for example, cooling and heating systems and plumbing (Grossi, Kunreuther (2005)). Therefore, the level of the entire building damage expected is given by the cumulative damage of all components, as illustrated in Figure 3. Ultimately, the output of this module provides vulnerability curves in terms of both the individual building component and the cumulative building components. Figure 3 shows the example of vulnerability curve against intensity of flood.

Another important thing to point out is the output of vulnerability module also includes critical estimates of uncertainty around expected damage value in terms of standard deviations. Together, the hazard and vulnerability modules comprise what it is known as a probabilistic risk analysis.

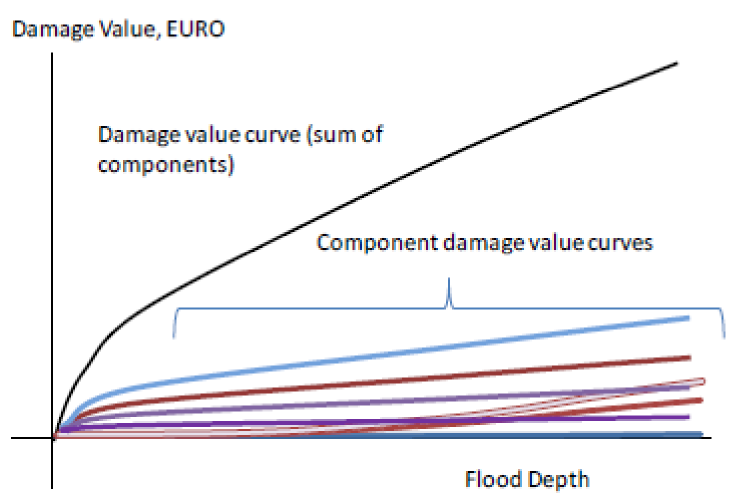

Figure 3 : The example of vulnerability curve (source: Parodi (2012))

Financial Module

This module concentrates on how to translate the estimates of physical damage to buildings and contents into estimates of monetary loss. This means the insured losses by applying insurance policy conditions to the total damage estimates together with the probability for relevant level of loss. This probability distribution of losses reveals the probability that any given level of loss will be surpasses in a given time period, for example in the coming year, as suggested in the annual rate of occurrence for each event in table 1. The output of this module is the Event Loss Tables (ELT) giving detailed information of event-by-event losses. Table 1 illustrates an example of a gross ELT with no insurance structure.

Table 1 : The example of Event Loss Table (Source: Parodi (2012))

Event	Rate	Expected Loss	Standard Deviation	Exposure
1 – Northern San Andreas 6.5	0.01	1,500,000	800,000	5,500,000
2 – Calaveras 6.5	0.01	3,000,000	2,000,000	15,000,000
3 – Hayward 7.0	0.02	6,500,000	5,000,000	50,000,000
4 – Northridge 6.5	0.03	8,000,000	6,000,000	90,000,000
5 – Southern San Andreas 5.0	0.03	10,000,000	7,000,000	95,000,000

Overall, catastrophe models aim to model the complex inherent components in catastrophe events through probabilistic risk analysis and conclude the likelihood and severity of catastrophe event. Obviously, this requires substantial amounts of data for model construction and validation and meanwhile is a collaboration job built by teams of highly-credentialed scientists and highly-trained structural engineers. Therefore, this leads to another issue about the uncertainties across different catastrophe models.

2.2 Catastrophe models and uncertainties

As observed in section 2.1, catastrophe models are a representation of complex physical phenomena using a probabilistic modelling approach, which itself contains various levels of uncertainty involved in catastrophe modelling process. Therefore, outcomes from different catastrophe models may result in quite different losses even when carrying out the analysis for the same or similar catastrophic events. As suggested by Parodi (2012), there are five types of uncertainties in actuarial practice. In my opinion, those uncertainties can also be applied to three module within catastrophe models.

Process uncertainty

Natural hazard itself happens without certainty therefore process uncertainty is the uncertainty that derives from dealing with inherently catastrophe event. This is intrinsic to the catastrophe event and cannot be reduced. . For example, even if we knew for sure how to measure the frequency of flood event through a Poisson distribution, we still not fully certain about when and where the flood would happen. The flood event itself could present random fluctuations from one year to the other, and these fluctuations would be driven exclusively by the natural hazard itself. How to measure the process uncertainty drives the motivation for building catastrophe modelling and that’s what the catastrophe modelling firms are doing until now.

Data uncertainty

A large issue in quantifying uncertainty within catastrophe models is the lack of data for characterising the three modules since it requires extensive amounts of data for catastrophe model construction and validation within each module [RMS, 2010]. The types of required data set can vary from detailed database of building inventories, data obtained from historical events, to detailed claims data and exposure data provided by clients, etc. In addition, for any model, the "garbage in, garbage out" concept holds irrespective of how advance or state-of-the-art a model may be [Parodi ,(2012)] . Therefore, the uncertainties regarding the data quality and accuracy before incorporating them into catastrophe models would be essential to recognize so that the underlying assumptions of different catastrophe models can lead to varieties in loss estimates and the uncertainty associated with these estimates accordingly.

Model uncertainty

Catastrophe models involve complex contents of modelling natural hazard, such as the understanding of scientific knowledge and cross-disciplinary approach between scientists and structural engineers, and no individual would claim for sure that the constructed model can perfectly reflect the catastrophe event. For example, the hazard module of catastrophe models requires simulating thousands of representative catastrophic events in time and space. And this job is built by a team of scientists, including geophysicists, climate scientists, seismologists, meteorologists and hydrologists, whose responsibility is to absorb latest scientific literature and assess the latest research findings to make sure that models incorporate the most recent scientific findings [AIR, 2012]. The vulnerability module requires estimating physical damage to various types of structures and their contents. This can be developed by a team of structural engineers, whose job is to incorporate published research, the results of laboratory testing, the findings from on-site damage surveys, as well as detailed claims data from insurer [AIR, 2012]. Therefore, different catastrophe modelling teams would incorporate different level of understanding of scientific knowledge into the catastrophe models resulting in variations in output together with the uncertainty around such output across different catastrophe models.

Parameter uncertainty

For each model, several parameters are never known with 100% accuracy, even if the model is correct. Such a concept also holds for catastrophe models. For example, if we use poison distribution to model the time of occurrence of earthquake in hazard module, the rate of occurrence still accompany with uncertainties which are hard to assess. Therefore, this also affects the variations in output across different catastrophe models.

Simulation uncertainty

Simulation uncertainty can be interpreted as “Approximation errors”, which means the errors derived not from a fundamental nature but from simply limitations of the methods used in modelling process [Parodi (2012)]. In the case of catastrophe modelling, most distributions are continuous but a simulation using a discrete distribution is presented for simplicity. For example, simulation techniques are used to sample the probability distribution of the level of structural damage (defined by none, minor, moderate, severe, or collapse) and approximation errors would occur in catastrophe modelling. Such errors also would affect the variations in calculation of the estimated loss across different catastrophe models.

Overall, the reliability of catastrophe models depends extensively on an understanding of the underlying physical mechanism which control the occurrence and behaviour of natural hazards [RMS, 2008]. There are various factors affecting the credibility of output from catastrophe models and no one individual would claim to have a complete understanding of all intricacies of these physical systems [RMS, 2008]. Therefore, teams of scientists and engineers have accumulated tremendous amounts of information and knowledge in catastrophe modelling area and indeed different catastrophe models may present various outcomes due to different level of understanding and interpreting those information and knowledge.

Now we need to look back at the purpose of catastrophe modelling, which is to assist insurers and reinsurers to anticipate the likelihood and severity of potential future catastrophes before they occur so that they can adequately prepare for their financial impact. However, the variation of outcomes across different catastrophe models would lead confusion to their users, to what extent, to believe the credibility of the results. Therefore, Guy Carpenter has introduced a framework of Model Suitability Analysis aiming to collaborate with their clients to develop their own views of risk management in catastrophe events.

3. CATASTROPHE MODEL SUITABILITY ANALYSIS

3.1 Introduction to Model Suitability Analysis (MSA)

Guy Carpenter has introduced MSA for the purpose of assisting its clients in the pursuit of their own view of risk, through a deeper understanding and a more sophisticated use of cat risk model results. It consists of eight components, each of which represents an analytical objective. These eight components are organised within three groups of tasks that aim at assessing the performance of catastrophe risk models (i.e. EVLUATION), their INTEGRATION into a particular risk view for the client, and the COMMUNICATION of finding to a client's internal and external audiences, including regulatory authorities [Franco, 2012]. Figure 4 shows details of this structure.

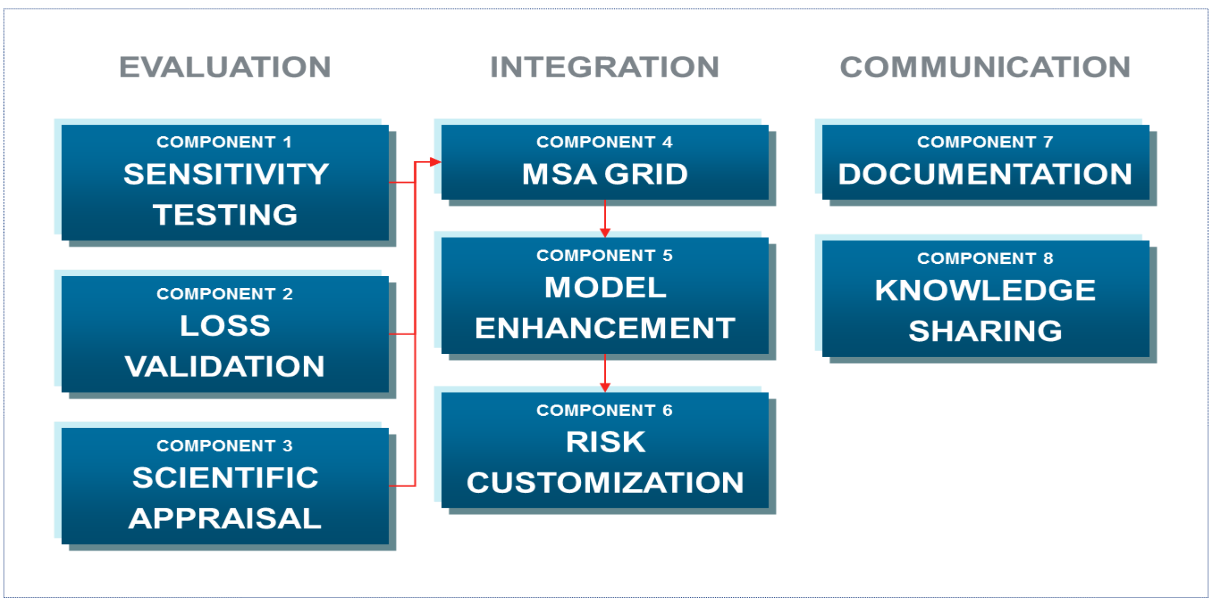

Figure 4: The framework of Model Suitability Analysis (source: Franco (2012))

MSA proposes a test-driven and client-specific evaluation of different catastrophe risk models, which is realised in the EVALUATION stage. It concentrates on assessing different catastrophe models driven by rigorously defined tests on a tailor-made basis, and summarizes all evaluation tests into the form of a “MSA grid”. The MSA Grid constitutes the foundation of risk customization, and is a key differentiator of the MSA process. As shown in Figure 4, all conclusions from the EVALUATION stage are summarized into the MSA Grid, which is a colour-coded table. The aim of this table is to provide insights into a client’s decision making process, regarding which catastrophe model would be most suitable to capture their risk characteristics. It also assists in pinpointing model traits that may constitute opportunities for model enhancement or adjustment and risk customization, ultimately leading to technical broking arguments that provide both brokers and clients an advantageous perspective for reinsurance placement. All these components lie within the second stage, called INTEGRATION. The last stage, referred to as COMMUNICATION**,** considers the most appropriate communication strategy, making available resources and training materials to demonstrate their view of risk to internal and external stakeholders, for example, regulatory authorities [Guy Carpenter, (2012)].

The following sections discuss the specific components within each stage in more detail.

<u>Sensitivity Testing (C1)</u>

This component focuses on identification of significant primary variables affecting loss results in order to understand how they affect catastrophe risk model performance and estimated losses. As discussed in section 2.2, there are hundreds of input parameters underpinning model hypotheses and assumptions. Examples of these are vulnerability region, inventory region and property characteristics, affecting catastrophe risk model results. Therefore, defining sensitivity tests to analyse the variation in loss results is helpful to understand catastrophe risk model results and their associated uncertainty. The findings from this component can also provide clients with insights into which type of portfolio data is advisable to collect, in order to reduce uncertainty associated to input exposure characteristics [Guy Carpenter, (2012)].

Loss Validation (C2)

This component aims at determining which catastrophe risk model can best capture the risk characteristics of a client's portfolio. This may be approached using tests that compare the modelled estimated losses with client's actual loss experience. The smaller the differences between modelled historical losses and a client's actual experience may indicate adequacy of claims data utilized by model vendors [Guy Carpenter, (2012)].

Scientific Appraisal (C3)

This component concentrates on evaluating the quality of key scientific assumptions that underlie catastrophe risk models. These scientific assumptions can play an important role in determining loss results; therefore it is necessary to set up an independent evaluation of those. Because of the complexity of scientific assumptions underlying catastrophe risk models, Guy Carpenter collaborates with external academic partners such as the Department of Civil Engineering and Engineering Mechanics at Columbia University for wind perils and the Istituto Universiario di Sudi Superiori in Pavia (Italy) for earthquake perils. The scientific appraisal consists of comparisons of hazard characteristics and assumptions within the catastrophe risk models and third-party datasets to provide insights into the suitability of catastrophe risk models [Franco, (2012)].

MSA Grid (C4)

MSA Grid serves as a summary of individual tests carried out in the EVALUATION process. It presents the summary in a colour-coded table, where each entry represents the performance of each catastrophe model for a specific testing criterion. This grid provides a simple way for clients to understand the catastrophe risk model's performance, while forming a basis for decision making, model enhancement and risk customization. Figure 5 provides an example of MSA Grid.

	C1: ST	C2: LV	C2: LV	C3: SA	C3: SA
	Relative RC Bldg Code	Klaus Loss Validation	Hi-Freq EP Validation	Agreement Dmg Funcs	Agreement Event Ftprts
MODEL 1	GOOD	10% ERROR	SO-SO MATCH	MATCHES RESEARCH	MATCHES UK MET
MODEL 2	GOOD	200% ERROR	POOR MATCH	NO MATCH	MATCHES UK MET
MODEL 3	NO LATEST AGE BAND	50% ERROR	SO-SO MATCH	MATCHES CLAIMS	NO MATCH UK MET

Figure 5 : The MSA Grid framework (source: Guy Carpenter (2012))

The headings (first two rows) represent individual tests from the components that belong to. The first column represents different catastrophe risk models available. The colour-coded area shows the result of a simple evaluation score for each test, where good performance corresponds to green, moderate performance to yellow and poor performance to red.

Model Enhancement & Risk Customization (C5&C6)

As discussed above, these two components are established on the basis of conclusions from the MSA grid, since it provides information on the model's suitability with respect to client's exposure. On these bases, MSA is able to develop the necessary adjustments of catastrophe risk models, and possibly blend them to best represent a particular client's risk profile.

Documentation & Knowledge Sharing (C7&C8)

These two components reflect MSA's communication strategy, responding to clients' motivation to communicate with internal and external stakeholders, such as risk managers and regulators. The documentation system produces documents in a standardised form that contains detailed conclusions for each defined tests within the MSA process. Clients are able to flexibly extract parts of these documents, and provide them as required to internal and external stakeholders.

In summary, MSA consists of a comprehensive process that contains all elementary components necessary for Cat model evaluation, Cat model integration and Cat model communication. The MSA Grid acts as a foundation for risk customization, and is a key differentiator of the MSA process since it contains significant information that allows identifying the most suitable catastrophe risk model. Exploration of methods to interpret the qualitative value in the MSA Grid to quantitative measurements is the essential aim of this paper. The relevant literature is reviewed in the next section.

3.2 Translating the MSA Grid into decision making

As shown in Figure 4, each entry of MSA Grid is interpreted through qualitative value by means of three different colours “green”, “yellow” and “red”, which reflects different level of both Guy Carpenter experts’ and its clients’ subjective judgments on catastrophe models as “good”, “moderate” and “poor”, respectively. A format of color-coded grids is clearly helpful for visualization purposes and has been appreciated by clients; however, the scheme is considered to be too qualitative to make a conclusion on the most suitable catastrophe risk model with a high degree of certainty. Hence, the main problem is how to quantify the degree of uncertainty inherent to subjective judgments, in order to determine the most suitable catastrophe risk model.

This section aims at exploring algorithms that dealt with uncertainty in experts’ subjective opinion. Cooke (1991) has suggested several models to estimate and quantify expert’s subjective opinions in decision making process in the field of science. Such a concept is appropriate for the interpretation of the MSA Grid, since each qualitative value within Grid cells may be regarded as expert’s subjective opinion, and the MSA Grid serves for decision making of the most suitable catastrophe risk model. To apply those models referred by Cooke (1991) to analyze the MSA Grid problem, we firstly need to discuss the problems of **subjective data** suggested by Cooke (1991) and compare them with the issues in interpreting qualitative value in the MSA Grid.

Firstly, the expert’s subjective opinions in science typically show extremely wide **spreads,** since the object of the opinion is usually a rare event, such as the average yearly probability of a core melt from a nuclear facility due to an earthquake [Cooke, (1991)]. However, MSA Grid contains only three levels of values of opinions, represented by three different colors. As a result, the spread of value represented by colors in the MSA grid seems not as wide as the spread of experts' subjective opinions typically found in science.

Secondly, experts' subjective opinions in science are not **independent**. That is, if an expert was a pessimist with respect to one judgment, there was a substantial tendency for him to be a pessimist on other judgments as well (Cooke (1991)). Such a problem may present itself in a different way in the MSA Grid. MSA Grid is collaboration between both catastrophe modeling experts and clients. That means the ultimate grids results from an agreement of all experts' subjective opinions. Nonetheless, there is one type of dependence concerning subjective data in the MSA Grid. There are various tests within each component e.g, C1, C2 and C3, several of which present relative dependences. That is, if one test was given an optimistic value there would be a tendency to be optimistic in respect of another correlated test.

Thirdly, subjective experts' opinions in science have a feature of **reproducibility**. That is, different experts applying the same risk assessment methodology to the same problem would obtain get similar results (Cooke (1991)). However, MSA Grid is a final output obtained on the basis of catastrophe modelers' collaborating approach together with client's view. As a result, it is accepted as a reliable product as long as clients agree with it, despite human judging errors may exist which are never known exactly. Therefore, this paper focuses on how to interpret the qualitative values in the MSA Grid, not on arguing how to obtain a reliable MSA Grid and therefore not tackling the problem of reproducibility in subjective data.

Fourthly, the question of whether the assessment of subjective experts' opinions in science is appropriate results in the problem of **calibration.** In this content, calibration is concerned with the extent to which assessed probabilities agree with observed relative frequencies (Cooke (1991)). Events studied in science are typically too rare to check for calibration and calibration requires a large number of available data. However, as discussed before, the MSA Grid problem in this paper is not concerned with checking whether the grid value is given appropriately, alternatively we are more concerned with interpreting qualitative value in the MSA Grid, therefore not tackling with the problem of lack of enough data to calibrate such uncertainty around subjective grid values.

The above four theoretical issues regarding subjective data make it challenging to apply experts' opinions' models directly to the MSA Grid issue. However, we can obtain some inspiration from the above discussion regarding subjective data and the theory of experts' subjective opinions in uncertainty. For example:

1. What is the spread of scores for each cell, and whether they should take a finite set of discrete values in a deterministic way or in a probabilistic way following with a certain distribution?
2. Dependence among judgments of different Cat modelers: are they optimistic or pessimistic?
3. The correlation between tests within the same component and correlation between tests across different components.
4. The relative significance across different tests, implying different weights to be assigned accordingly.
5. The subjective opinions provided on the basis of the collaboration approach associated to the MSA Grid may imply uncertainty, since there

will always be uncertainty in relation to whether a particular Cat model's results is suitable for a client's portfolio.

The above inspirations enable me to generate the necessary assumptions for exploration of appropriate methods to interpret qualitative value in given MSA Grid to quantitatively measure different catastrophe risk models, ultimately determining the most suitable model. The relevant methods and assumptions are described in detail in the next section.

4. MODEL SCORING IN THE MSA GRID

4.1 Testing example of the MSA Grid and assumptions

This section focuses on scoring different catastrophe risk models on the basis of given MSA Grid. With consideration of the audience of MSA Grid are Guy Carpenter 's clients, one should develop simple and accessible methods to interpret qualitative values into quantitative measurements in order to achieve effective communication on the most suitable model with the clients. Hence, this section presents two methods and provides a testing example of MSA Grid in terms of deterministic and stochastic modeling.

To simplify the scoring methodology, it is necessary to illustrate the underlying assumptions. Firstly, each cell in the MSA Grid can be seen as subjective data, which intrinsically contains a degree of uncertainty. Hence, the value in MSA Grid can be assumed with discrete figure in deterministic modeling, and the uncertainty around it can be reflected by assumed probability distribution in stochastic modeling. Moreover, the level of correlation across different tests and the dependence of subjective judgments into different test's results can result in complexity of scoring method. Therefore, no correlation between different tests' results is assumed and independent given subjective judgments are assumed. Furthermore, various tests may inherent difference in significance according to clients' exposure portfolio and their preference. In practice, the significance of individual test's result depends on both the impact of aggregate losses and client's preference. However for simplicity, the assumption of same significant level across various tests is necessary.

The summary of underlying assumptions in the following testing example is

- Each cell of MSA grid has taken up a finite set of discrete value among {1, 2, 3}, where 1, 2, 3 represents Poor, Moderate, and Good respectively.
- The uncertainty around each cell of MSA Grid can be represented by an assumed probability distribution
- All tests' results are independent to each other.
- All tests have the same weights of significance.

Based on above assumptions, Table 2 demonstrates a theoretical testing example of MSA Grid consisting of two available catastrophe models and 12 tests, and each test has the same weight of significance.

Table 2 : Testing example of the MSA Grid

	C1 : Sensitivity Tests				C2: Loss Validation				C3: Scientific Appraisal			
	C1-1	C1-2	C1-3	C1-4	C2-1	C2-2	C2-3	C2-4	C3-1	C3-2	C3-3	C3-4
Model 1	3	1	3	2	1	3	3	3	2	3	2	1
Model 2	3	3	2	3	1	1	2	2	2	1	2	2
weight (exposure)	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

4.2 Scoring method 1- aggregation by excluding the best N tests' scores

4.2.1 Method 1's inspiration

In financial mathematics and financial risk management, there are several risk measures. Variance[1] and standard deviation[2] are traditional techniques. In

[1] In definition, variance is a measure of how far a set of numbers is spread out [Wikipedia]
[2] In definition, standard deviation shows how much variation or dispersion from the average(mean, also called expected value) exists.[Wikipedia]

addition, VaR[3]and TVaR[4] are both widely used to measure the risk of loss on a specific portfolio of financial assets. For example, if a portfolio of stocks has a one-day 5% VaR of $1 million, that means there is a 0.05 probability that the portfolio will fall in value by more than $1 million over a one day period if there is no trading [Wikipedia]. If a portfolio of stocks has a one-day 5% TVaR of $ 1 million, it means the average value of the portfolio falling in more than 5% VaR would be $1 million.

The above example shows that TVaR focuses on measuring the average of the worst scenarios. Furthermore, if taking simulation methodology of VaR and TVaR as another example, the idea can be emphasized. If one has 100,000 simulated scenarios, all equally likely, one would calculate the 99th percentile of simulated scenarios as the estimate of 99% VaR. To calculate 99% TVaR, one can sort the 100,000 simulated scenarios firstly and then taking the worst 1,000 scenarios only, and averaging the amounts of those scenarios.

To score a series of tests' scores in MSA Grid, one can focus on the trend of the worst scenarios, in this context, which means excluding the best test score among a sorted series of tests' scores. The trend can be investigated by taking off the best tests' score one by one and aggregating weighted average scores of the rest tests. Thus, the pattern of aggregation scores of each model, calculated by weighted average scores among available tests, gives an idea of the more appropriate model: the higher scoring of the model, the better model would be.

---

[3] Called Value at Risk, in definition, VaR is a threshold value such that the probability that the loss on the portfolio over the given time horizon exceeds this value is the given probability level [Wikipedia]

[4] Called Tail value at risk, in definition, TVaR is the average value beyond a certain VaR and can be regarded as a conditional expected value. [Wikipedia]

Therefore, scoring method 1 gives rise to plot aggregation scores of each model against the number of best test excluded together with calculation for basic statistics, such as **Mean, Standard Deviation, Weighted Average, Weighted Standard Deviation and Media** for the purpose of comparing competing models.

Suppose the score for each cell in table 2 is sorted from smallest to largest, expressed as X_j, j=1 to 12. The weight of significance of each test can be expressed asω_j, j=1 to 12. In the case of same weight for each test, ω_j= 0.0833, j=1 to 12. Then the summary of formulas regarding to basic statistics in our example is as follows:

- Mean $\bar{X}$= $\sum_{j=1}^{12} Xj/12$
- Standard Deviation $s = \sqrt{\frac{\sum(X_j-\bar{X})^2}{11}}$
- Weighted Average $\bar{X}_\omega = \sum_{j=1}^{12} \omega_j Xj$
- Weighted Standard Deviation= $sd_w = \sqrt{\frac{\sum_{i=1}^{12} \omega_i (X_i-\bar{X}_\omega)^2}{\frac{(N'-1)\sum_{i=1}^{12}\omega_i}{N'}}}$ where $N'$is the number of non-zero weights.
- Median= the value separating the higher half of the tests' score within MSA Gird
- Weighted Average Scores of remaining tests after excluding the best n tests' scores

  = $\frac{\sum_{i=1}^{j} \omega_i \times X_i}{\sum_{i=1}^{j} \omega i}$ j=12, 11, 10…

As assuming the same weight of significance for each test, the statistics of weighted average or weighted standard deviation should be equal to mean and standard deviation respectively. It is only necessary to compare Mean, Standard Deviation and Median between two models.

4.2.2. Deterministic approach

In the case of deterministic approach, it means that modeling scoring of both models ignores the level of subjective uncertainty in value given in a MSA Grid. In other words, the subjective value is given with one hundred percent certainty.

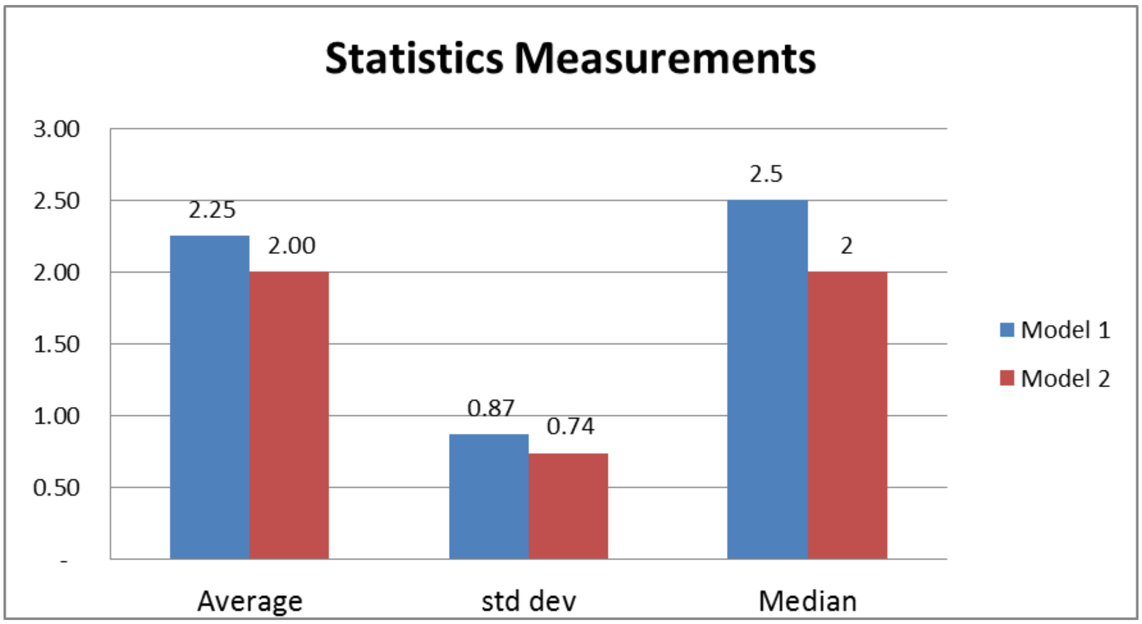

Figure 6 : comparison of statistics measurements between MODEL 1 & 2

Figure 6 shows the result of comparison of basic statistics measurements between MODEL 1 and 2. MODEL 1 has higher average scoring and higher median, which indicates it containing more “good” tests than MODEL 2. In addition, higher standard deviation in MODEL 1 indicats that the spread of tests’ scores deviating from mean score of MODEL 1 is wider than that of MODEL 2, implying more “moderate” tests within MODEL 2. Thus, comparison of basic statistics measurements indicates that MODEL 1 is better than MODEL 2.

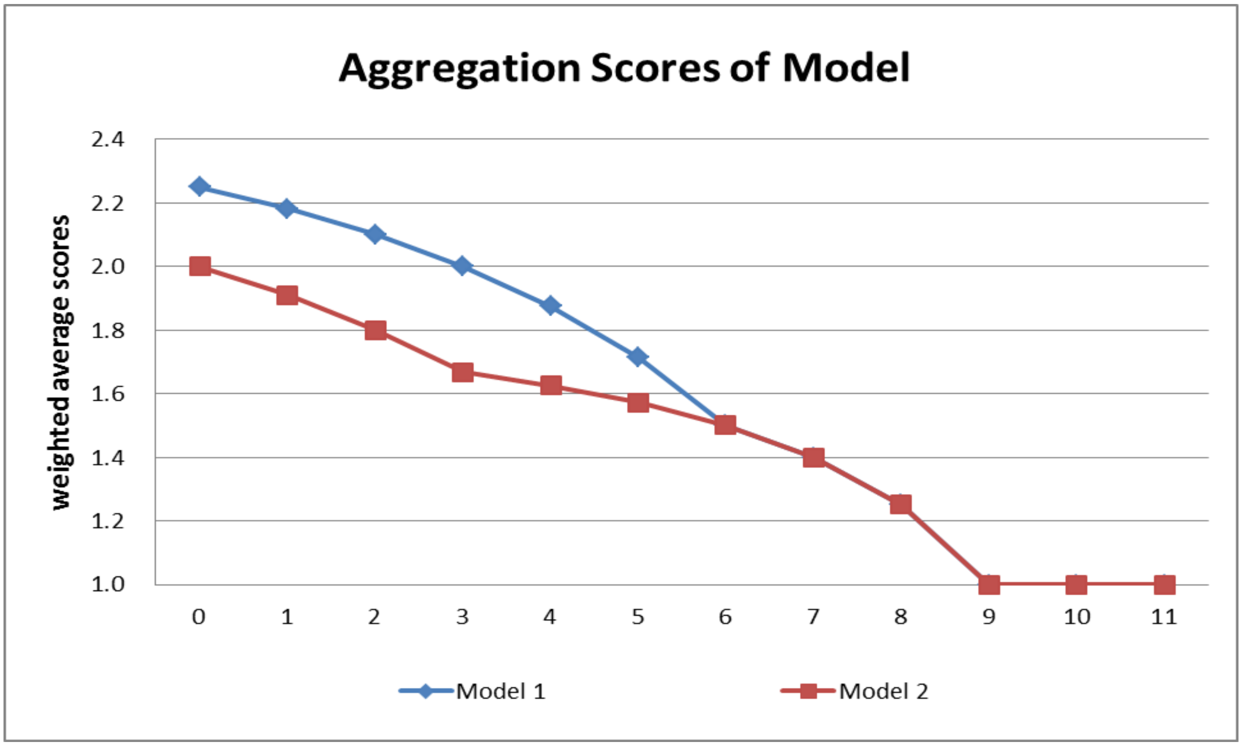

Figure 7: Comparison of weighted average scores of MODEL 1 & 2

Figure 7 confirms the results obtained from figure 6 and also concludes that MODEL 1 performs better than MODEL 2. Because the weighted average scores of tests of MODEL 1 has remained higher than that of MODEL 2 until the exclusion of the sixth best score test and afterwards remained the same with each other. This indicates that MODEL 1 has relative advantage of obtaining higher weighted average scores even excluding the best five scoring tests. The slope of curve also provides simple criteria of identifying the better model: the quicker the curve decreasing indicates that the overall score of each model is driven by a few "good" tests.

4.2.3. Stochastic approach

In the case of stochastic approach, the score of each test can be regarded as subjective data intrinsically containing a degree of uncertainty. This means, even if an expert judges the test score of "Good" and the actual score of the test should be assigned with "Good", there still exists the probability for the same expert giving judgment of "moderate" and "poor". Such uncertainty can be interpreted by assigning a probability distribution to each value of test score and sample the probability distribution through simulation in order to capture the feature of uncertainty in subjective judgment. The assumption of associated probability distribution assigning to each test score in testing example of MSA Grid (Table 2) is illustrated in table 3 as below.

Table 3: Example of assumption of probability distribution

Judgement	test score	probability distribution		
		Good=3	Moderate=2	Poor=1
Good	3	0.75	0.20	0.05
Moderate	2	0.30	0.35	0.30
Poor	1	0.05	0.20	0.75

One can simulate 1000 series of test scores for each model on the basis of distribution assumption in table 3. In this case, one can observe the average value of statistics measurements together with their standard errors based on 1000 simulated scores, and compare the trend of curve of mean aggregation scores between each model.

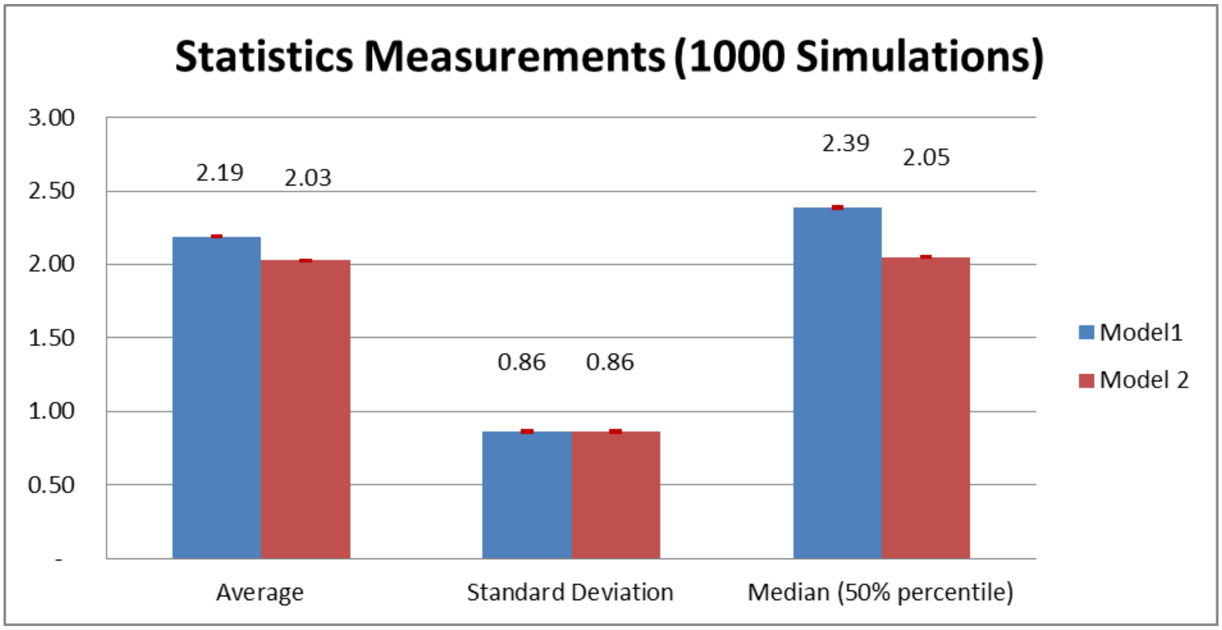

Figure 8 : Comparison of statistics measurements between model 1 & 2 (1000 simulations)

After considering uncertainty in MSA Grid, Figure 8 shows that MODEL 1 has slightly lower mean and median than itself in deterministic analysis but the opposite situation applies to MODEL 2 (Figure 6). However, MODEL 1 still has higher average value of mean and median than model 2 based on 1000 simulation results. In addition, the standard deviation of both models shows little difference. Thus, basic statistics measurements in a stochastic approach also indicate that MODEL 1 is better than MODEL 2.

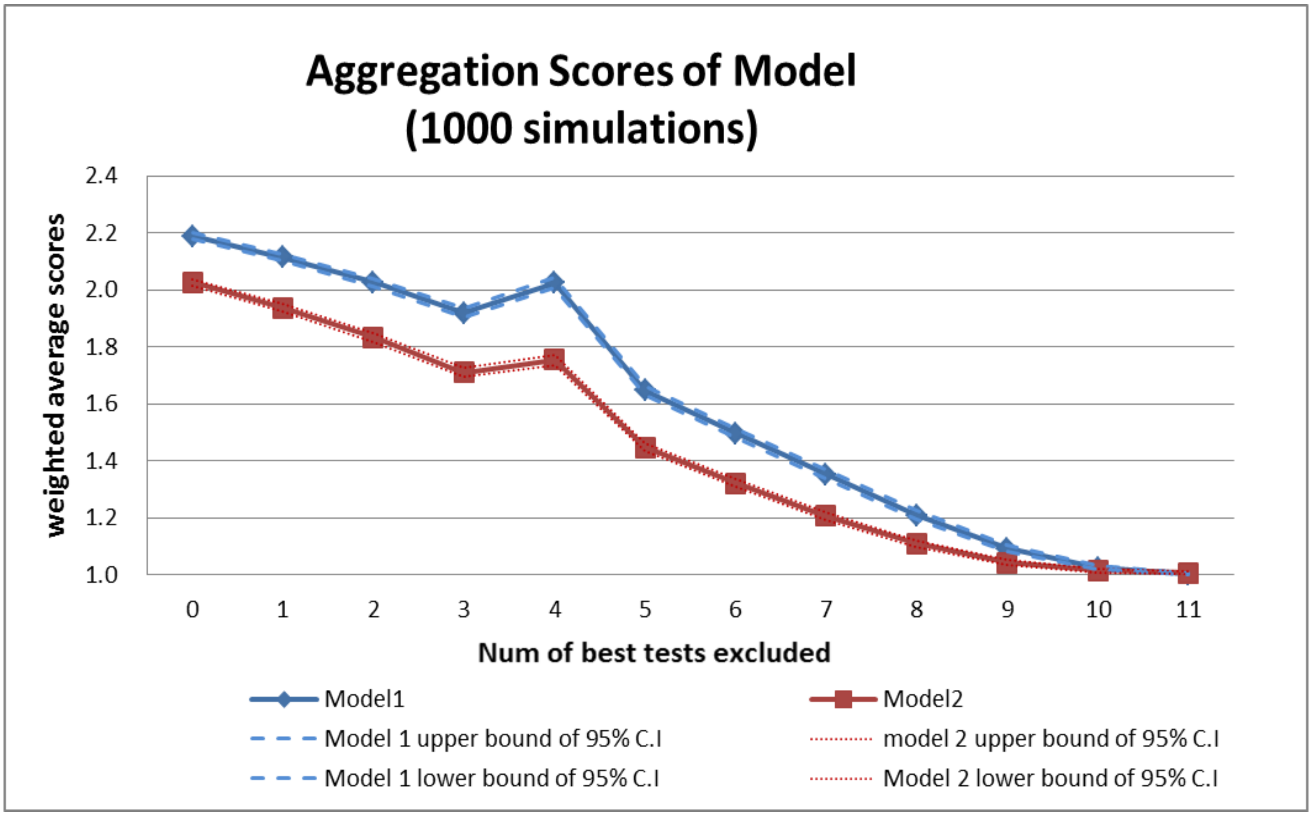

Figure 9: Comparison of weighted average scores of model 1 & 2 against the number of best tests excluded (1000 simulations)

Figure 9 shows the average aggregation scores for MODEL 1 and 2 on a sample of 1000 simulations plus their corresponding bounds of 95% confidence interval for the mean of each model. The distance between upper and lower bounds of 95% confidence interval of mean aggregation score for both models are close enough to make conclusion with great level of precision. The average aggregation scores of MODEL 1 has remained higher than that of MODEL 2 until the exclusion of the last best score test, thus Model 1 performs better even considering the underlying assumption of uncertainty. In a stochastic approach when deriving the suitability of models, one must consider the appropriateness of probability assumption for uncertainty.

4.3 Scoring method 2 – focusing on “Good” and “Poor”

4.3.1 Method 2’s inspiration

The main idea deriving this method is the bias feature of subjective data. This means that people always make subjective judgments for “Good” and “Poor” with more certainty than that for “Moderate”. Hence, the assessment only focuses on the number of tests’ results of “Good” and “Poor” excluding that of “Moderate” when comparing suitability between two models. One can plot a figure of x axis representing weighted number of “Poor” tests and y axis representing weighted number of “good” tests. So the observation of “Good” against “Poor” for each model can be scatter plotted on the above figure. Taking the slope of line connecting the origin with the observation value as a quantity measure, one can differentiate two models on the basis of criteria: the higher the slope, the better model represents.

4.3.2 Deterministic approach

Without considering uncertainty of subjective judgment, one can plot the ideas discussed in section 4.3.1 in Figure 10. The plot of MODEL 1 falls in the area above the diagonal of y=x where is exactly the plot of MODEL 2 falls along. This can conclude that both MODEL 1 and 2 are acceptable since the slope of lines connecting the origin to the observation value of MODEL 1 and 2 are equal and greater than 1. But relatively speaking, MODEL 1 performs better than MODEL 2 since blue line (MODEL 1) has higher slope than red line (MODEL 2).

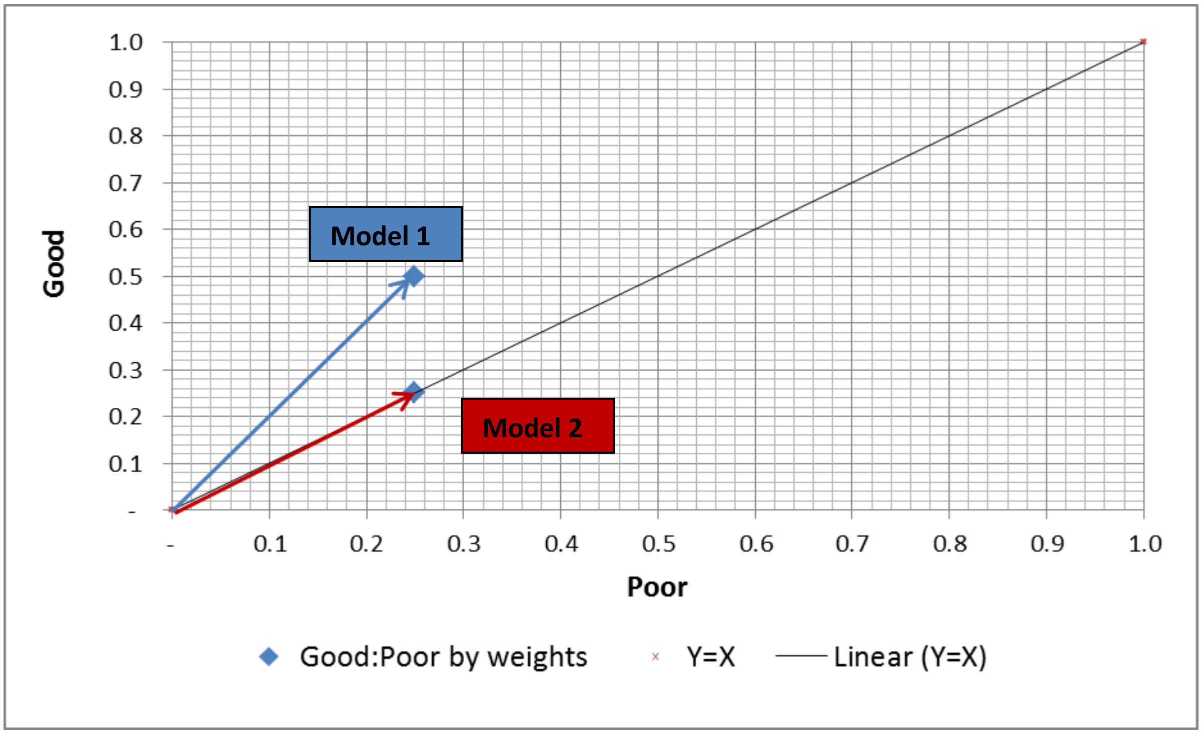

Figure 10 : Plot of the weighted number of tests "Good" against "Poor" for MODEL 1 & 2

4.3.3 Stochastic approach

Considering the degree of uncertainty given in subjective judgments and applying the same assumption of subjective uncertainty in Table 3, one can plot the similar figure of weighted number of "Good" against "Poor" for each model shown as Figure 11 below.

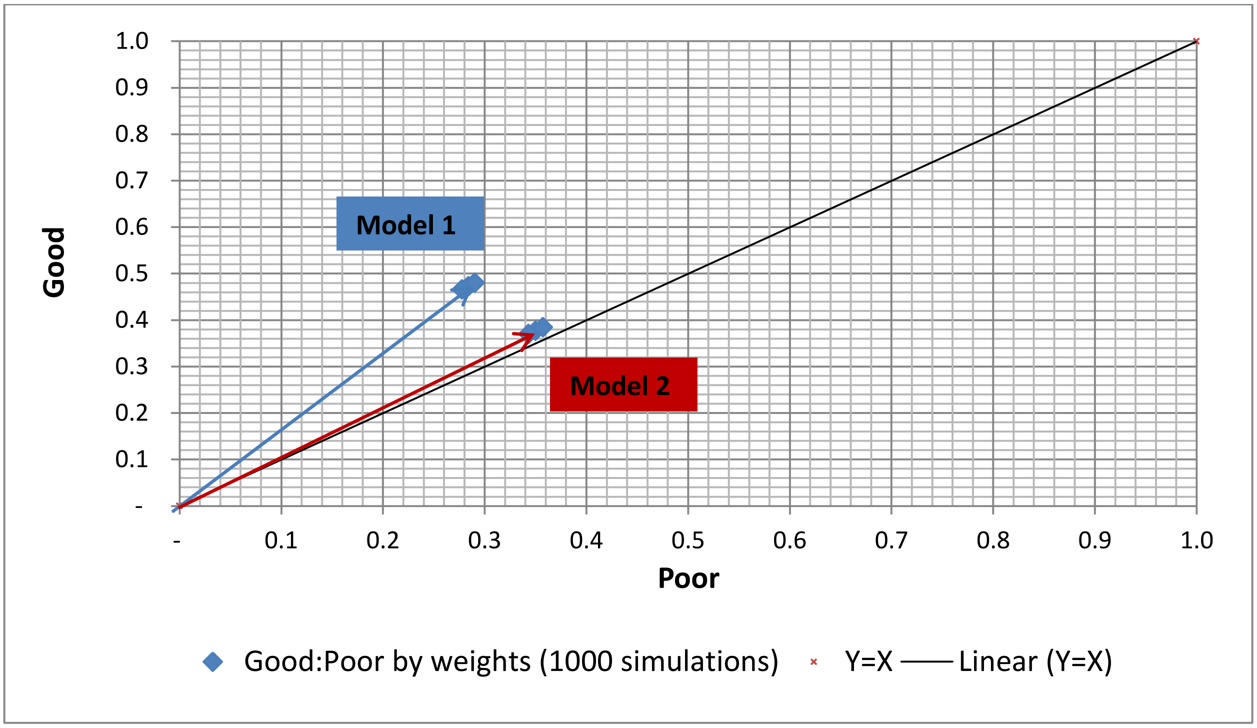

Figure 11: Plot of the weighted number of tests "Good" against "Poor" for MODEL 1 & 2 (1000 simulations)

Figure 11 shows the average weighted number of tests "Good" against "Poor" for MODEL 1 and 2 on a sample of 1000 simulations plus their corresponding bounds of 95% confidence interval for the mean measurement of each model. Visually speaking, the upper and lower bounds of 95% confidence interval of mean weighted number of tests for each model are overlapping distributed around the mean value. This can conclude that the average weighted number of tests "Good" over "Poor" on 1000 simulations can result in conclusion with precision. The conclusion confirms with the deterministic approach that MODEL 1 performs relatively better than MODEL 2 however both models are acceptable.

To view the distribution of weighted number of “Good” tests against “Poor” tests among 1000 simulations, one can scatter plot the simulated values for each model shown as Figure 12 and 13 below.

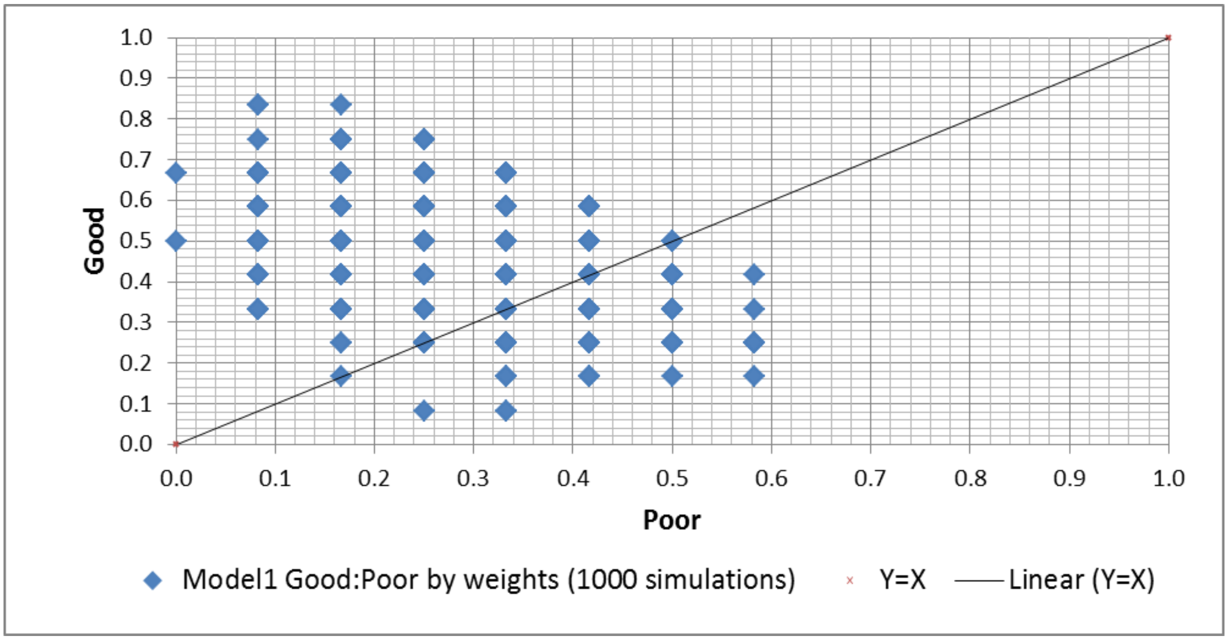

Figure 12: Scatter plot of weighted number of tests "Good" against "Poor" for MODEL 1

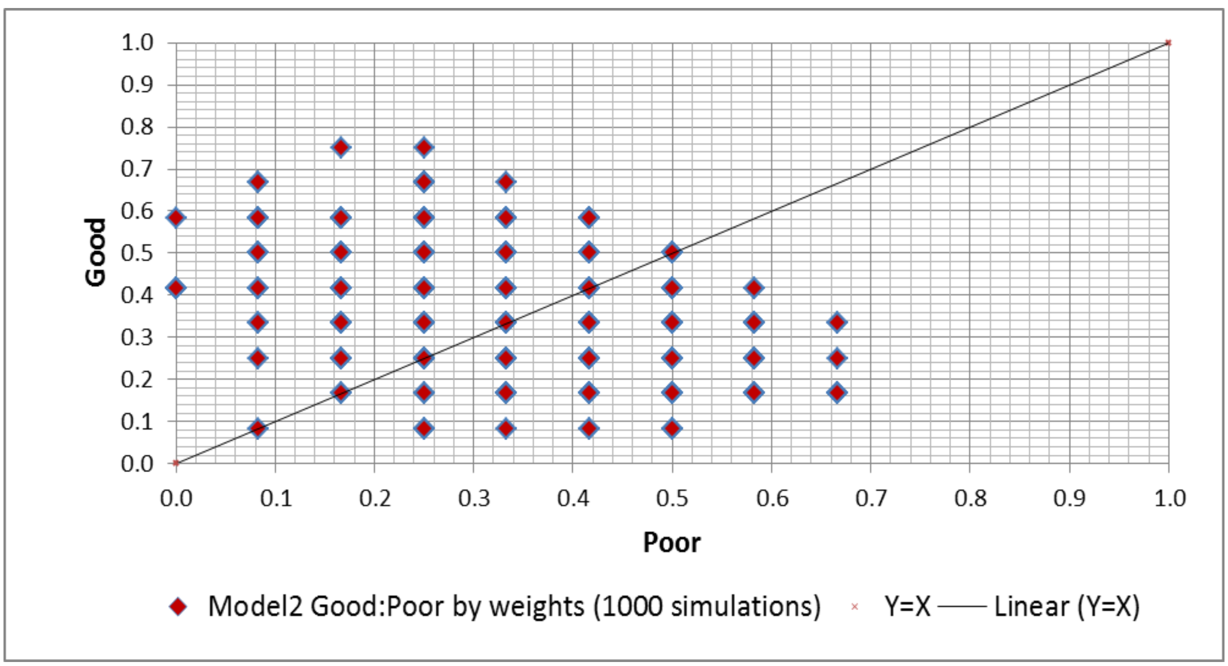

Figure 13: Scatter plot of weighted number of tests "Good" against "Poor" for MODEL 2

Comparing Figure 12 and 13, one can observe that the number of data points lying on and above the diagonal for each model is greater than those below the diagonal. This indicates that both MODEL 1 and 2 contain more “Good” tests than “Poor” tests with higher possibility under the consideration of subjective uncertainty. But relatively speaking, MODEL 1 presents more data points above the diagonal than MODEL 2.

In summary, both scoring methods in terms of deterministic and probabilistic view can conclude that MODEL 1 performs better than MODEL 2 in the Grid example (Table 2) although MODEL 2 is also acceptable if one only focuses on the “Good “and “Poor” tests. Both scoring methods are simple in application especially when communicating with clients and particularly concentrates on relative comparison between two models because in practice clients usually would like to have an idea of which model would be more suitable for their exposure portfolio. Thus, next section will present a case study in application of the above two scoring methods in practice.

5. CASE STUDY

5.1 Introduction to case study

In this section, earthquake lossess in Turkey provide an interesting example as a case study to apply scoring methods 1 and 2 described in section 4, in order to differentiate competing catastrophe risk models. Turkey is located in a seismically active area, known as the Anatolian Block which is sandwiched between the Arabian , African and Eurasian plates [AIR,2009]. In history, Turkey has experienced several significant seismic events since 20th century, especially two successive adverse earthquakes in 1999, İzmit earthquake with magnitude 7.5 on August 17 and Düzce earthquake with magnitude 7.2 on November 12. Those earthquakes caused more than 19,000 fatalities, 48,000 injuries and the displacement of approximately half a million people [AIR, 2009]. It is estimated that the earthquakes have caused more than 1.4 billion euro in insured losses and some 14 billion euro in total damage in 1999 currency [AIR, 2009]. Insurers and reinsurers require comprehensive and sophisticated catastrophe modeling tools to help them fully understand the scale of the risk they face in such seismically active areas, and to develop effective strategies to manage the potential losses from such a high-impact catastrophe event [AIR, 2009].

Guy Carpenter has studied earthquakes in Turkey from a risk management perspective, and incorporates this case study into the Model Suitability Analysis (MSA) framework. The case study concentrates on test 3 in component 3 scientific appraisal within MSA. This entails comparison of the seismicity rates for all the combined seismogenic zones of Turkey, as per two competing catastrophe risk models (AIR & RMS), and rates obtained from scientific research by Grunthal et al.(2010) [Guy Carpenter, (2013)]. The scientific appraisal component (test C3-3) results in plots of the annual exceedance rate of earthquake occurrence against the earthquake magnitude, referred to as Gutenberg–Richter distribution, for each seismic zone of Turkey. Figure 14 displays an example of a plot, derived from a particular catastrophe risk model for Turkey.

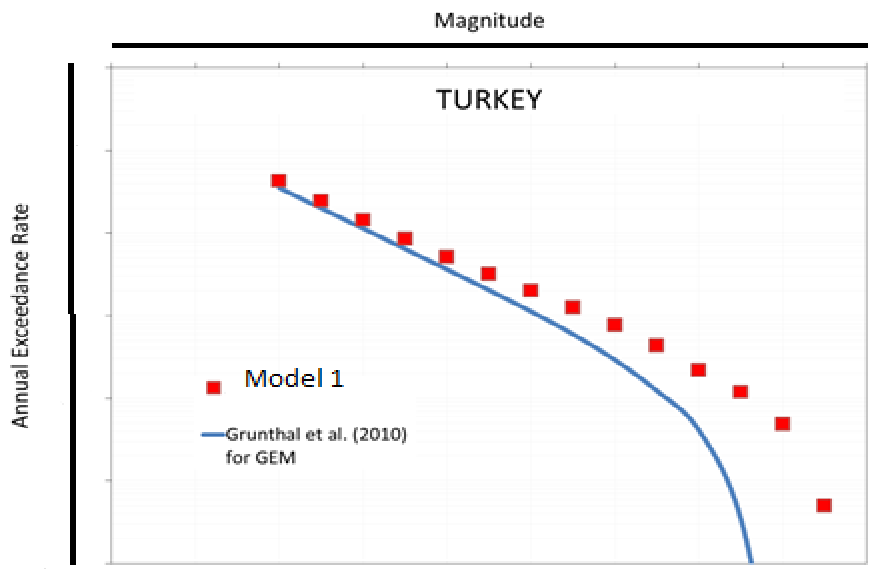

Figure 14: Plot of Annual Exceedance Rate of earthquake against Magnitude in Turkey

From the observation of each plot zone by zone, catastrophe modeling experts can evaluate how close the modeled seismicity rates and the predicted rates by Grunthal et al area, which can subsequently be summarized into a MSA Grid zone by zone. One can apply the scoring method 1 and 2 discussed in section 4 to build this MSA Grid, and conclude from it which catastrophe model would produce the closest results to the scientific results by Grunthal et al., provided his results are reliable. Another potential application in this case study is to explore the assignment of weights to each seismic zone in MSA Grid according to the exposure of a particular insured portfolio in the corresponding zones.

5.2 How to create a MSA Grid and associated weights

The MSA Grid is commonly created on the basis of subjective and fuzzy judgments of experts. The ideal path to develop a MSA Grid, however, would be on the basis of objective algorithms, superseding the subjective element to the largest possible extent. However, as discussed in section 2.2, the uncertainty associated to catastrophe model results makes it complex deriving an appropriate algorithm to solve this problem. Thus, how to create the MSA Grid is out of the scope of this case study. This section only focuses on how to differentiate competing catastrophe models on the basis of a given MSA Grid, created by catastrophe modeling experts in Guy Carpenter in this case.

To create the MSA Grid of test C3-3, catastrophe experts firstly divide the map of Turkey into several seismogenic zones, which represent similar levels of seismic hazard. Figure 15 shows the distribution of seismogenic zones in Turkey.

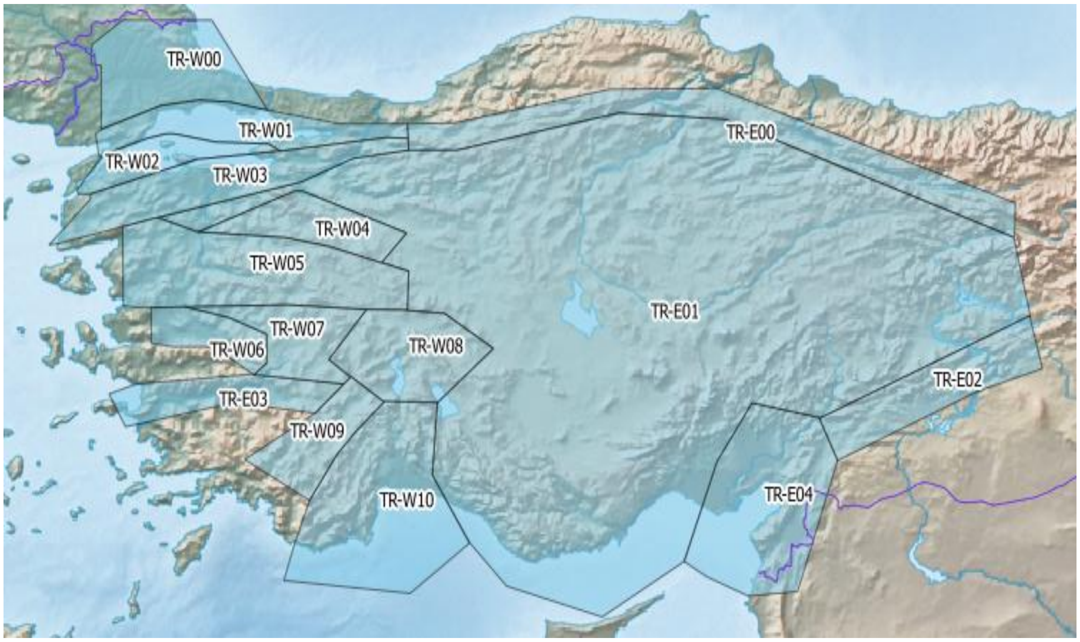

Figure 15: distribution of seismogenic zones of Turkey (Source: Guy Carpenter (2013))

Secondly, plots can be derived from competing catastrophe risk models, zone by zone. These may be compared to results from scientific research. Figures 16 and 17 show the Gutenberg–Richter distribution plots corresponding to two catastrophe models, zone by zone.

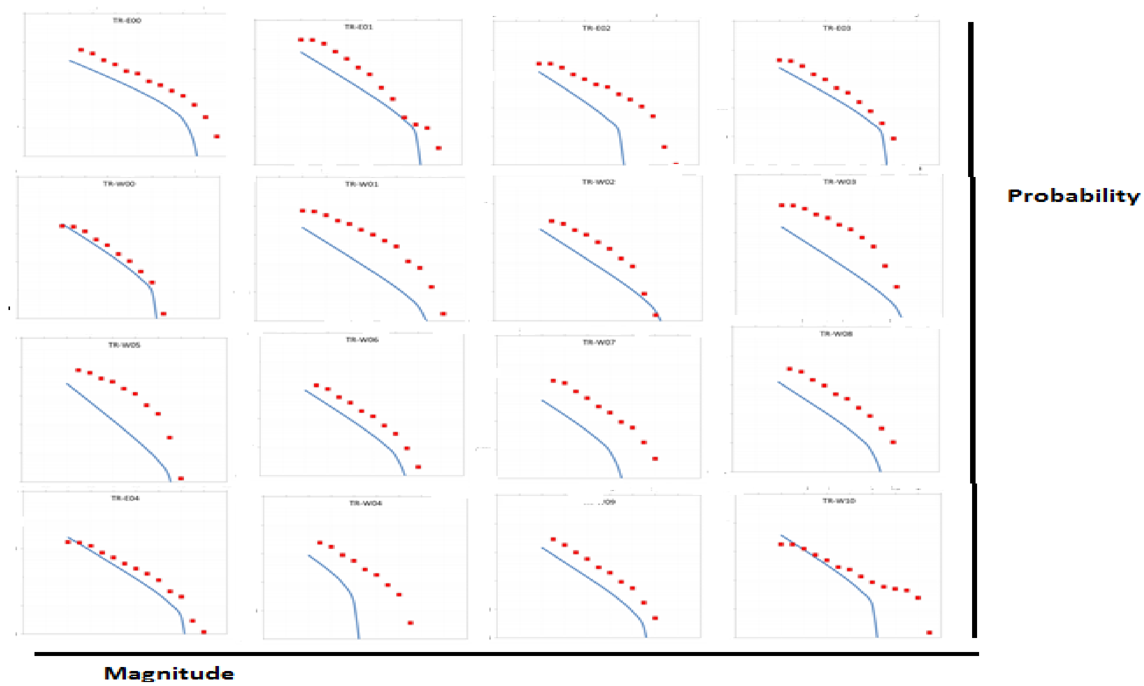

Figure 16: Gutenberg–Richter distribution zone by zone of Model 1 (Source: Guy Carpenter (2013))

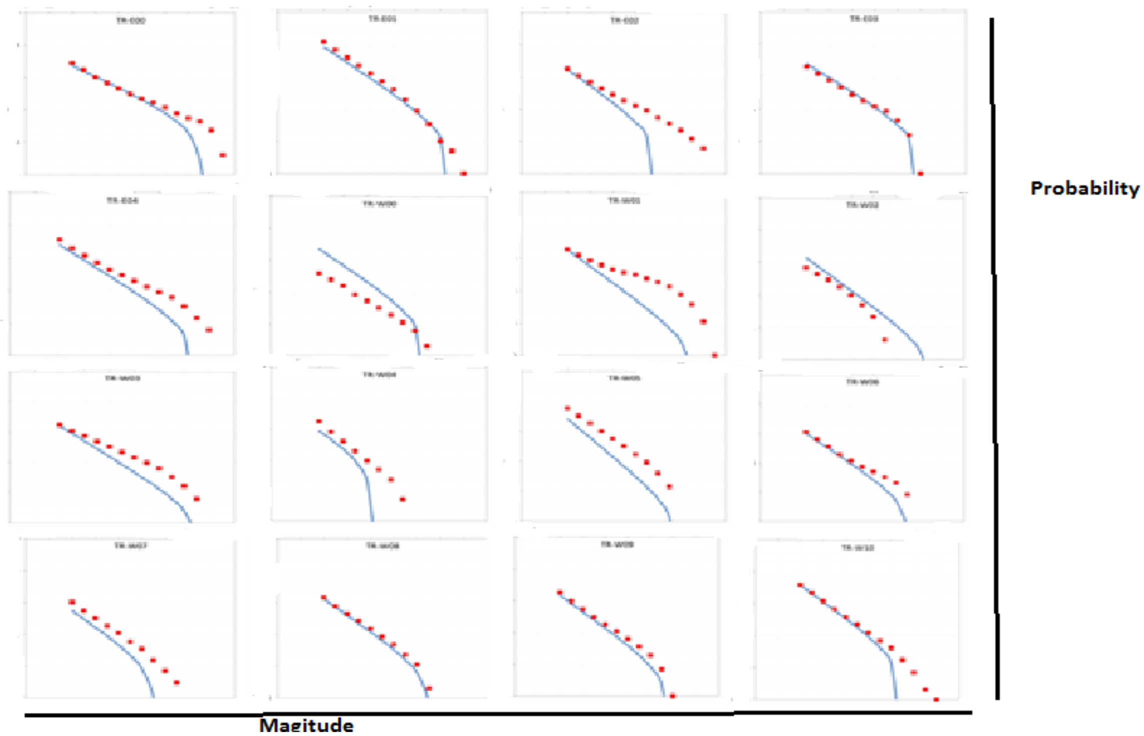

Figure 17 : Gutenberg–Richter distribution zone by zone of Model 2 (Source: Guy Carpenter (2013))

Thirdly, the MSA Grid of test C3-3 can be developed from Figure 16 and 17 by considering catastrophe modeling experts' subjective judgments, as illustrated in Table 4. The MSA Grid therefore summarizes the assessment of seismicity rates zone by zone in Turkey, as per the comparison between catastrophe models' views and that of scientific researchers.

Table 4 : Test C3-3 MSA Grid of Turkey earthquake

Zone	TR-E00	TR-E01	TR-E02	TR-E03	TR-E04	TR-W00	TR-W01	TR-W02	TR-W03	TR-W04	TR-W05	TR-W06	TR-W07	TR-W08	TR-W09	TR-W10
Model 1	3	3	1	3	2	2	1	3	2	2	2	3	2	3	3	3
Model 2	1	2	1	2	3	3	1	2	1	1	1	2	1	1	2	2

The next step focuses on how to determine the weights to be associated to tests' results. In practice, a client's exposure portfolio can be used to differentiate seismogenic zones in terms of total insured value in each zone. Different clients would have different distributions of total insured value in each zone; therefore different conclusions regarding to model suitability may be derived under the same MSA Grid and the same scoring method, depending on portfolio of the client. Thus this case study concentrates on evaluating the effect of weights assigned to different tests on scoring model suitability. Tables 5, 6 and 7 show the distribution of exposure zone by zone for clients 0, 1 and 2 respectively.

Table 5: Distribution of Total Insured Value for seismogenic zone by zone in Turkey (Client 0)

Zone	TR-E00	TR-E01	TR-E02	TR-E03	TR-E04	TR-W00	TR-W01	TR-W02	TR-W03	TR-W04	TR-W05	TR-W06	TR-W07	TR-W08	TR-W09	TR-W10
client 0 (weight)	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%

Table 6 : Distribution of Total Insured Value for seismogenic zone by zone in Turkey (Client 1)

Zone	TR-E00	TR-E01	TR-E02	TR-E03	TR-E04	TR-W00	TR-W01	TR-W02	TR-W03	TR-W04	TR-W05	TR-W06	TR-W07	TR-W08	TR-W09	TR-W10
Client 1 (weight)	1%	22%	1%	3%	4%	13%	35%	1%	12%	2%	1%	3%	1%	0%	1%	3%

Table 7 : Distribution of Total Insured Value for seismogenic zone by zone in Turkey (Client 2)

Zone	TR-E00	TR-E01	TR-E02	TR-E03	TR-E04	TR-W00	TR-W01	TR-W02	TR-W03	TR-W04	TR-W05	TR-W06	TR-W07	TR-W08	TR-W09	TR-W10
Client 2 (weight)	1%	13%	1%	3%	35%	22%	4%	1%	12%	2%	1%	3%	1%	0%	1%	3%

Client 0 represents equal weighted exposure portfolio, which is created in theory for the purpose of comparison. However, Client 1 and 2 represents the practical exposure zone by zone. When comparing the distribution of exposure between Client 1 and 2, we can observe that they have the same exposure for all zones except zone TR-E01, TR-E04, TR-W00 and TR-W03 , which have been color coded by yellow in Table 7. In addition, one can observe that Client 1 concentrates insured values in zone TR-W01 and TR-E01, being different from those of Client 2:

zone TR-E04 and TR-W00. Not surprisingly, Client 2 has a majority business in the zones where MODEL 2 has greater value in the MSA Grid, while Client 1 has a majority business in the zones where MODEL 1 has greater value in the MSA Grids. Therefore, application of scoring methods 1 and 2 to the MSA Grid of Turkey earthquake in respect of different clients' exposure may provide interesting insights, which may give rise to business assessment of decision making on which catastrophe risk model would agree with the independent scientific prediction, so implying the most suitable catastrophe model to rely on.

5.3 Scoring catastrophe risk models for Turkey Earthquake

This section concentrates on applying scoring methods 1 and 2, described in section 4, to Test C3-3 MSA Grid of Turkey Earthquake in respect of different clients' exposure portfolios for each seismogenic zone.

5.3.1 Scoring method 1 and conclusion

Deterministic Approach

Figure 18 : comparison of statistics measurements between MODEL 1 & 2

Figure 18 shows the weighted average and weighted standard deviation of MODEL 1 and MODEL 2 for considered clients. For clients 0 and 1 MODEL 1 has higher weighted average scores; however for Client 2 MODEL 2 has higher weighted average scores. This indicates that Client 2 has insured more business in the zones where MODEL 2 performs better than MODEL 1. Client 2 witnesses a significant difference in weighted standard deviation between MODEL 1 and 2. This indicates that the majority of exposure portfolio of Client 2 is lying in the region where MODEL 2 observes different value from MODEL 1 in given MSA Grid. In this case, Client 2 has the majority of exposure in region TR-E04 (35%) where the MSA Grid has greater value in MODEL 2. Overall, one may draw the conclusion that for Client 1 MODEL 1 performs better but for Client 2 MODEL 2 is more suitable.

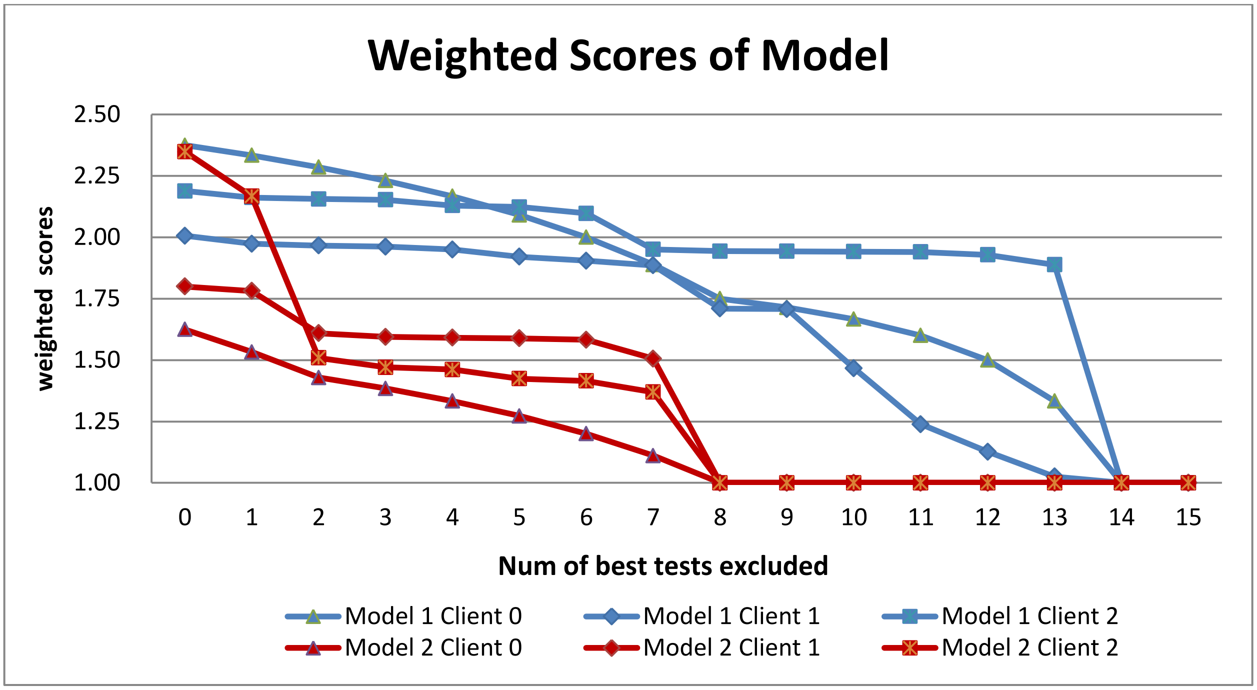

Figure 19 : Comparison of weighted average scores of MODEL 1 & 2 against the number of best tests excluded

Figure 19 shows the pattern of weighted average scores of MODEL 1 and 2 across different clients' exposure portfolios. For equally weighted exposure (client 0), MODEL 1 shows absolute advantage over MODEL 2. As for Client 1, MODEL 1 presents higher weighted average scores in respect of MODEL 2, although the advantage of MODEL 1 is less significant than in the case of equal weights. There is an interesting finding in the case of Client 2 that MODEL 2 has higher weighted average scores until exclusion of the best two scoring tests, but afterwards MODEL 1 has higher weighted average scores when excluding the rest of best scoring tests. This is mainly because MODEL 2 only performs better than MODEL 1 for the most heavily weighted zones for Client 2, so after excluding the best two scoring tests, MODEL 1 has higher weighted average scores. As a result, for Client 1, one can conclude that MODEL 1 is better than

MODEL 2. However, for Client 2, it is difficult to conclude that MODEL 2 is better overall. But if Client 2 only concentrates on model's performance among its majority exposure regions, one can conclude that MODEL 2 is better. Thus, under consideration of weights across different tests, scoring method 1 shows a heavy dependence on the distribution pattern of weights to associated tests.

Stochastic Approach

Applying the probability distribution shown in table 3 to this case study, one can generate the mean statistic measurements and mean aggregation scores of each model for Clients 1 and 2, on the basis of 1000 simulation data points. Client 0 has been created only for comparison purposes in the deterministic approach so it will not to be discussed in this section.

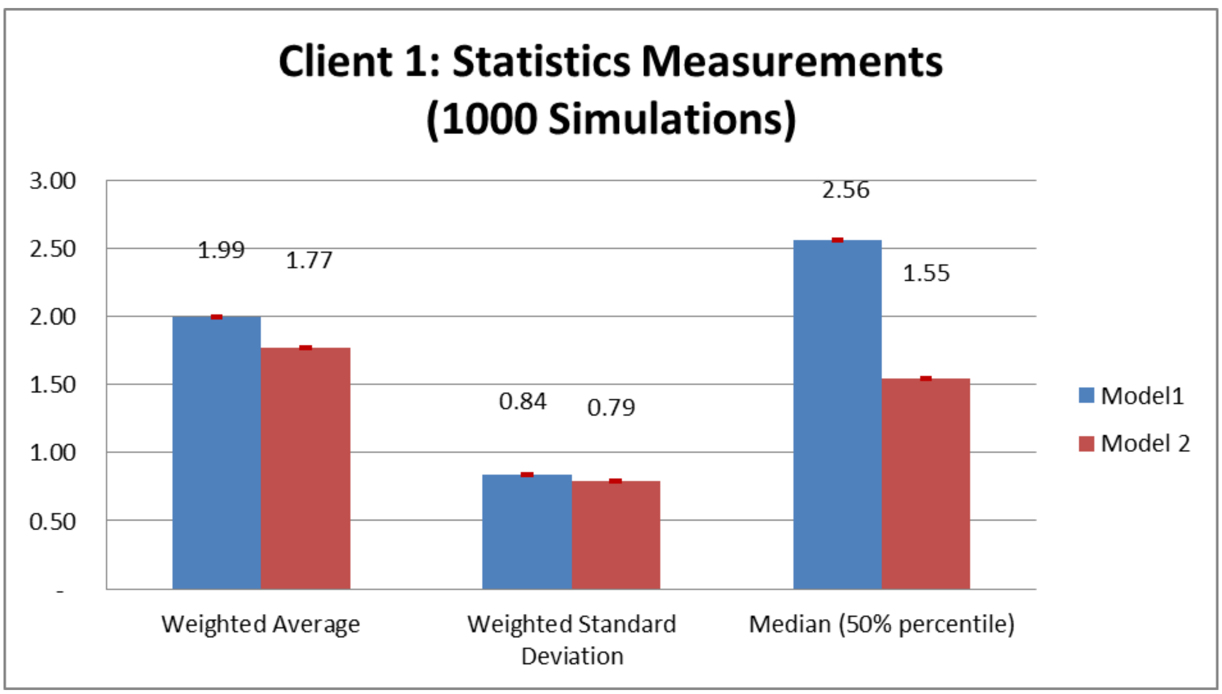

Figure 20 : Comparison of statistics measurements between MODEL 1 & 2 (Client 1)

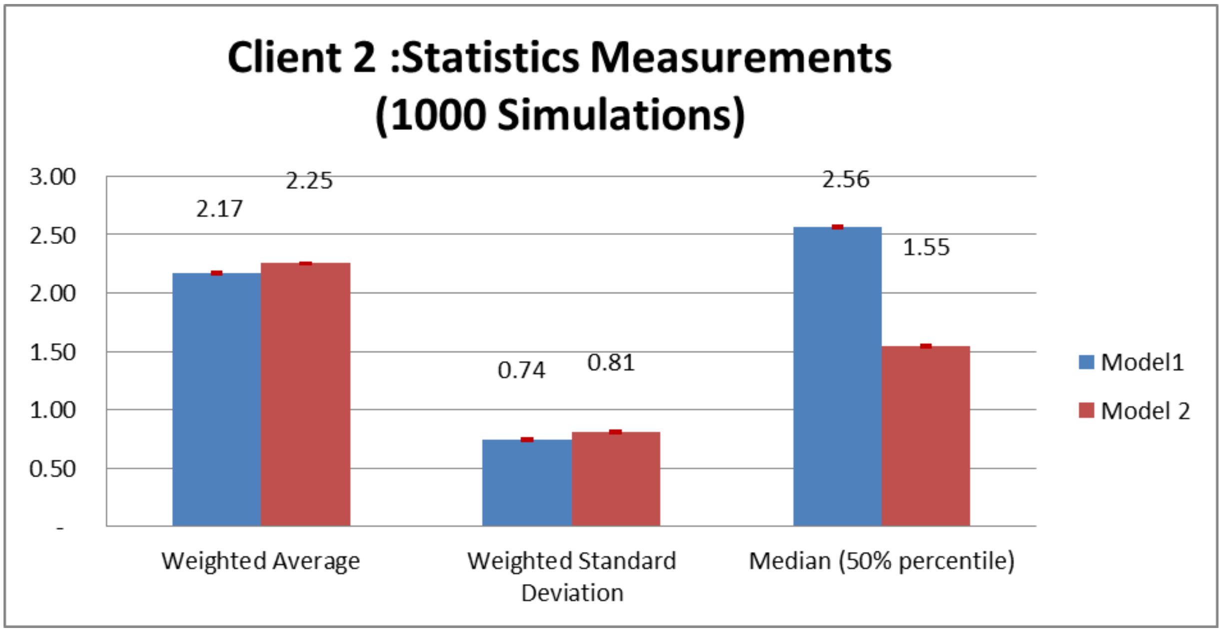

Figure 21 : Comparison of statistics measurements between MODEL 1 & 2 (Client 2)

Figures 20 and 21 show the mean of statistics measurements corresponding to 1000 simulation data points, i.e. weighted average, weighted standard deviation and median, and the corresponding standard errors of mean values between MODELS 1 and 2 for Clients 1 and 2 respectively. For Client 1, MODEL 1 has absolutely higher weighted average than MODEL 2 however for Client 2 MODEL 2 has slightly higher value than MODEL 1. Client 2 has higher weighted average scores than Client 1 for both models, which agrees with the previous finding that Client 2 has more insured exposure in the zones where MODEL 2 scores higher than MODEL 1. In the case of stochastic approach, the difference in weighted standard deviation between MODEL 1 and 2 for Client 2 is not significant as in deterministic approach. Both clients have the same median for both models, implying that the median only depends on the MSA Grids, regardless of the weights distribution. Thus, one can conclude that MODEL 1 performs better than

MODEL 2 for Client 1, and that MODEL 2 is more suitable than MODEL 1 for Client 2 on the basis of the assessments of statistics measurements.

However, if we explore the trend of weighted average scores when excluding the best-scoring tests one by one separately for Client 1 and 2, further significant observations are made. Figures 22 and 23 shows the mean value of weighted average scores for MODEL 1 and 2 corresponding to 1000 simulation data points, together with their corresponding 95% confidence intervals for Client 1 and 2 respectively.

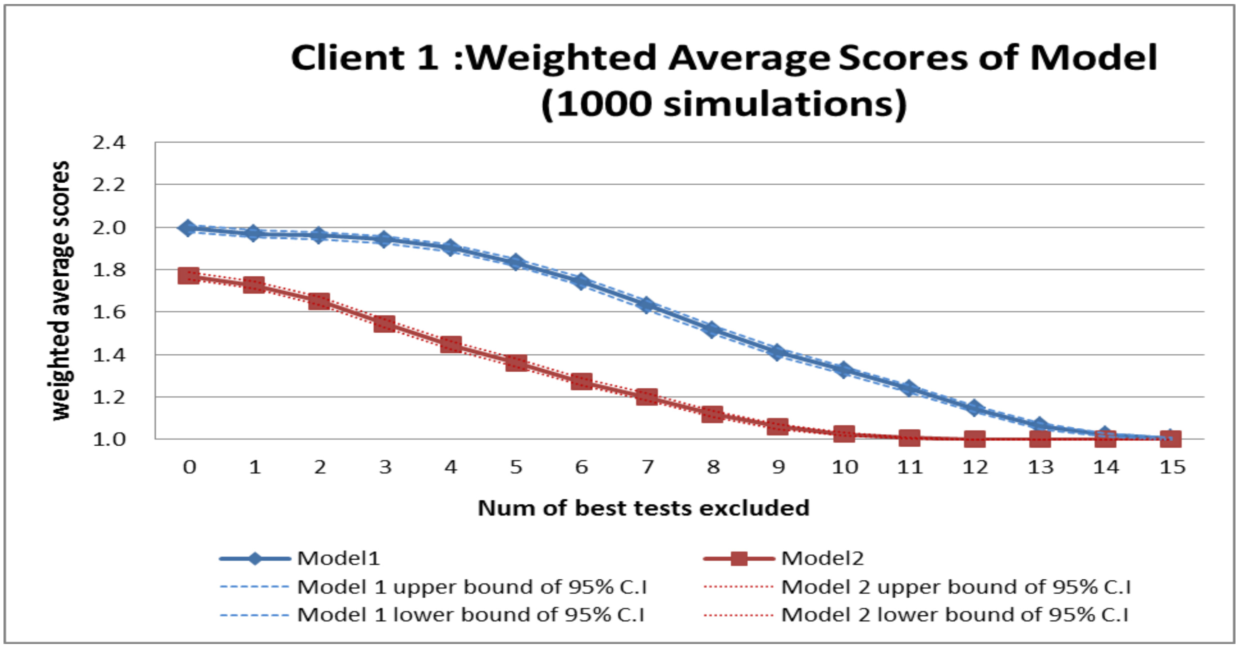

Figure 22 : Comparison of weighted average scores of MODEL 1 & 2 for Client 1

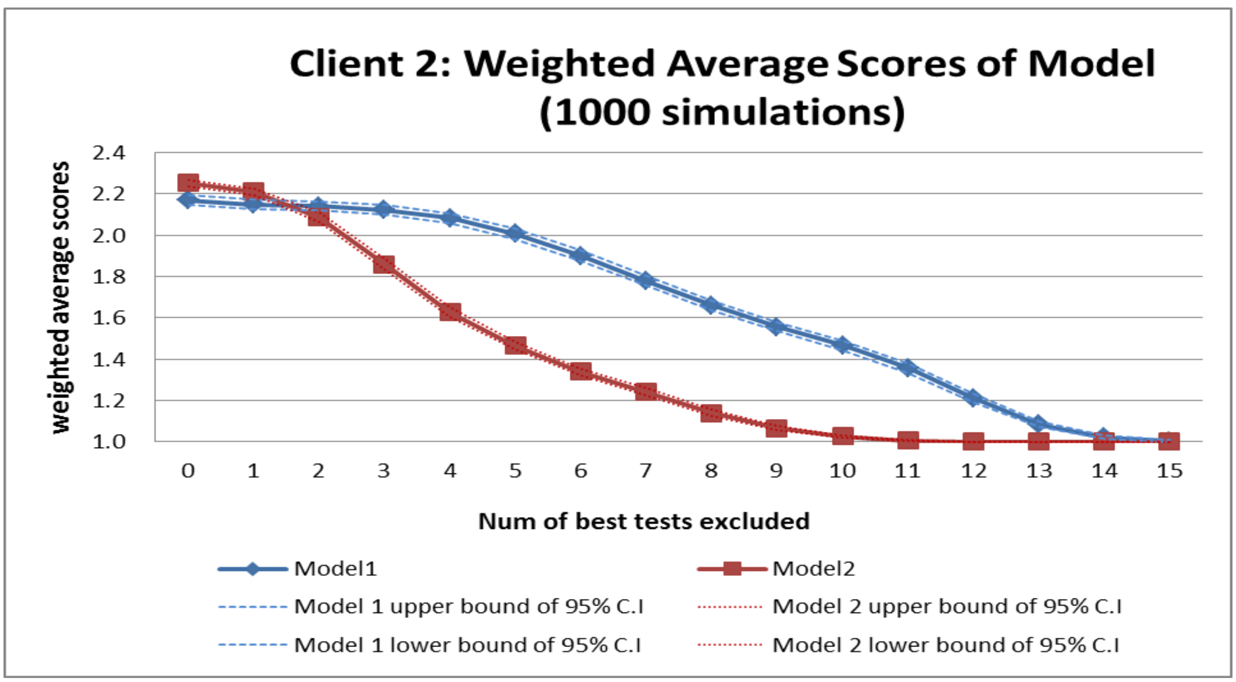

Figure 23 : Comparison of weighted average scores of MODEL 1 & 2 for Client 2

When excluding the best-scoring tests one by one for Client 1, the weighted average scores of MODEL 1 have remained higher than those of MODEL 2, until the point at which the last best test's score is excluded. Thus, it may be concluded that for Client 1 MODEL 1 performs better than MODEL 2.

As for Client 2, results become more complex to interpret. MODEL 2 has higher weighted average scores than MODEL 1 before excluding the best two scoring tests, however after excluding further best-scoring tests, the curve of weighted average scores for MODEL 2 drops significantly lower than that of MODEL 1. This indicates that the overall weighted average scores for MODEL 2 are driven by a few "Good "tests.

This is reinforced by comparing the decreasing slope of curves for both models and both clients, which shows that MODEL 2 is more sensitive to the distribution of weights than MODEL 1. This is due to the fact that the decreasing slope of the curve of MODEL 2's weighted average scores for Client 2 drops more rapidly than that of Client 1. Thus, it is difficult to draw a conclusion for Client 2 that MODEL 2 performs better than MODEL 1, although the measurement of statistics shows that MODEL 2 is better. This agrees with the results of the deterministic approach.

5.3.2 Scoring method 2 and conclusion

<u>Deterministic Approach</u>

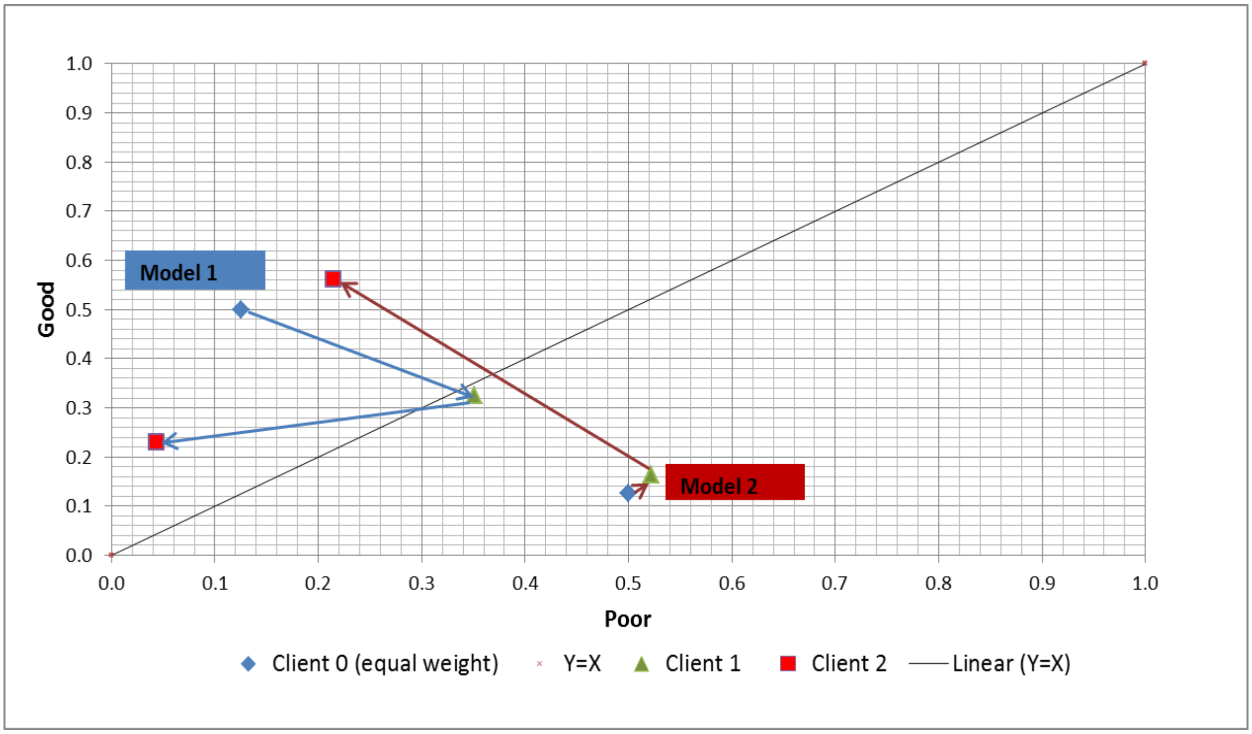

Figure 24 : Plot of the weighted number of "Good" tests against "Poor" tests for each model for Clients 0, 1 and 2

Figure 24 describes explicitly the movement of position on weighted number of "Good" tests against "Poor" tests from Client 0 (equal weighted exposure), Client 1 to Client 2. Under equally weighted exposure, MODEL 1 has absolute advantage over MODEL 2. Because MODEL 2 is located below the line Y=X, it may be judged as a poor model. After applying the exposure portfolio of Client 1, MODEL 1 moves towards a location below the diagonal Y=X, indicating that MODEL 1 is not as suitable for Client 1 as MODEL 2 is. However, while applying the exposure portfolio of Client 2, the position of both models completely changes. Both MODELS 1 and 2 move to the area above the diagonal, which may be regarded as acceptable. This means that both MODELS 1 and 2 are suitable for Client 2, but MODEL 1 is more suitable because it is located closer to the vertical line “Good” (Y axis) . Therefore, one may conclude that scoring method 2 is quite sensitive to the corresponding weights attributable to “Good” and “Poor” tests across different clients.

Stochastic Approach

Applying the probability distribution shown in table 3 to scoring method 2, one can generate the mean value of the weighted number of “Good” tests against “Poor” ,on the basis of 1000 simulations across Clients 0, 1 and 2.

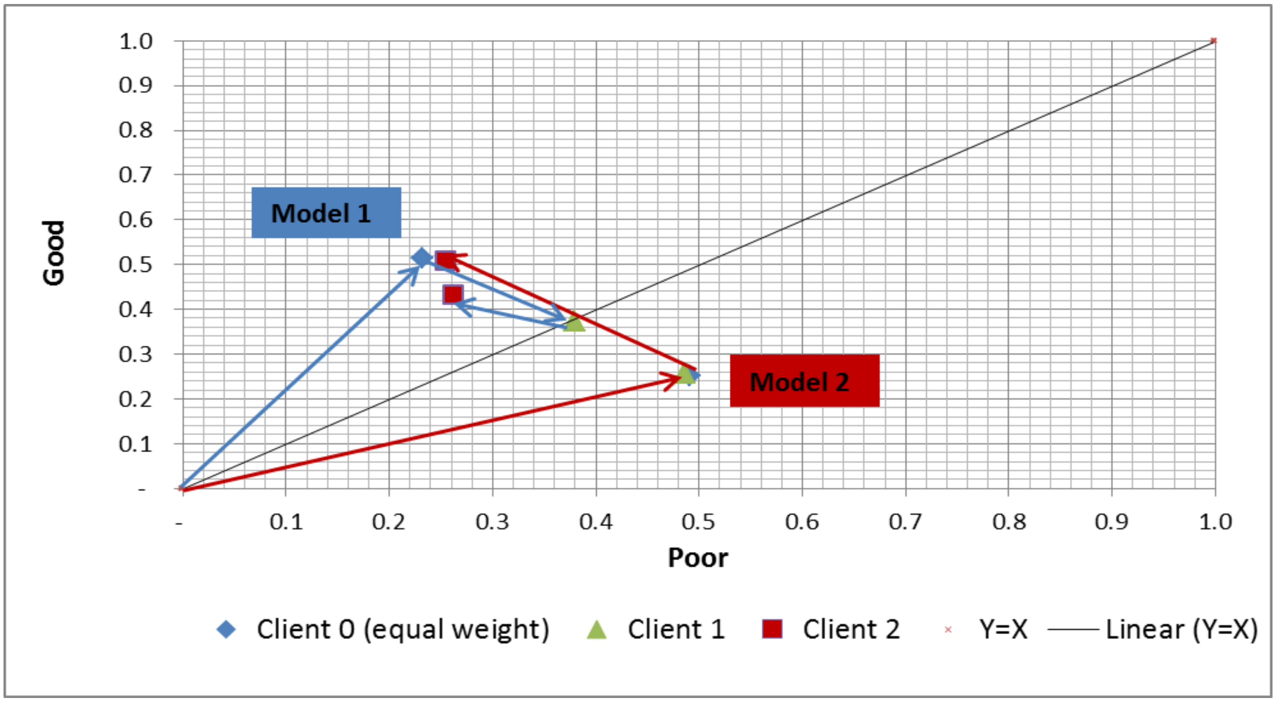

Figure 25 : Plot of the mean of weighted number of “Good” tests against “Poor” tests for each model across clients 0, 1 and 2 (1000 simulations)

Figure 25 describes the movement in the position of weighted number of “Good” tests against “Poor” tests for Client 0 (equal weighted exposure), Client 1 to Client 2, with consideration of subjective uncertainty. Figure 25 shows very different trends of movement for the considered clients, as compared to Figure 24. By comparing the movement of Client 0’s position between the deterministic and the stochastic approaches, one may conclude that sampling subjective uncertainty has a significant impact on the position of weighted number of

“Good” tests against “Poor” tests plot for scoring method 2 even, even without changing the exposure portfolio of the considered clients. This is due to the fact that the position of Client 0 on the plot has moved a lot for both models, before and after considering the subjective uncertainty. As for Client 1, both models may be considered to be poor, but MODEL 1 performs relatively better than MODEL 2. This agrees with the results of the analysis that do not consider subjective uncertainty. However, Client 2’s position in the plot for both models changes in opposite directions, before and after considering the subjective uncertainty. This means that before applying uncertainty distribution, MODEL 1 performs better than MODEL 2, but after its consideration, the opposite can be seen. This shift indicates that Client 2’s exposure portfolio is more sensitive to the uncertainty distribution, due to the fact that more insured exposure lies within the few “Good” tests in MODEL 2, as compared to MODEL 1.

To view the distribution of weighted number of “Good” tests against “Poor” tests among 1000 simulations for MODELS 1 and 2, one can scatter plot the simulated values for each model shown as Figures 26 to 29 as below.

Comparing Figure 26 and 27, one may conclude that MODEL 1 performs better than MODEL 2 for Client 1, since MODEL 1 has more data points scattered across the area above the diagonal. This agrees with the observation from the deterministic analysis.

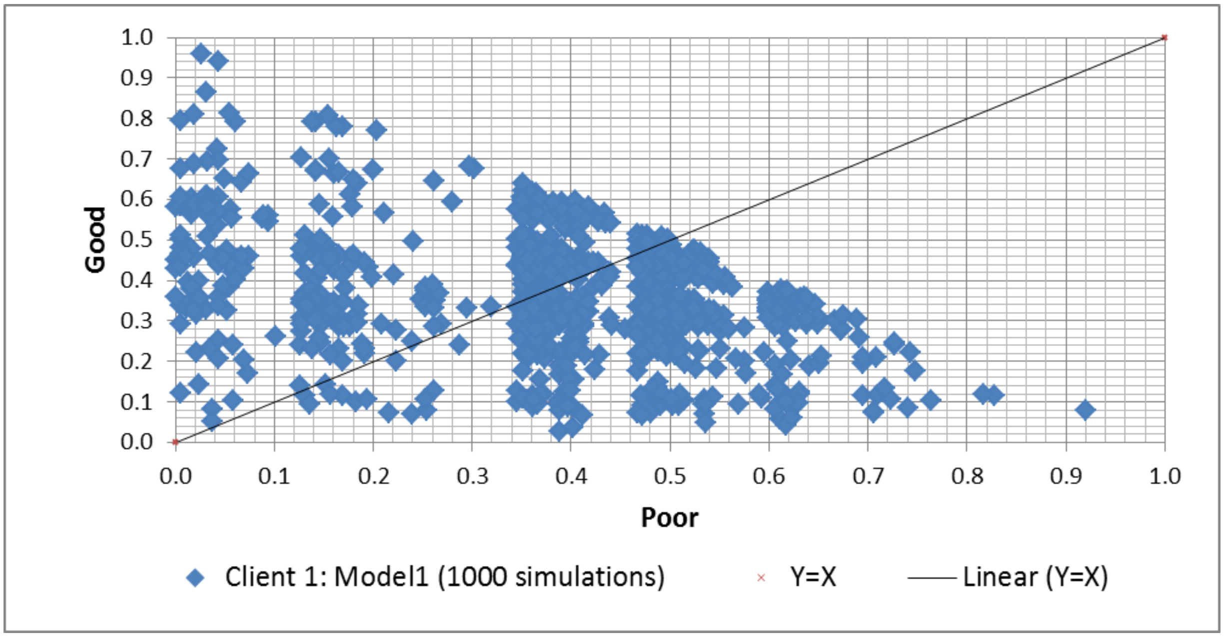

Figure 26 : Plot of weighted number of "Good" tests against "Poor" tests of MODEL 1 for Client 1 (1000 simulations)

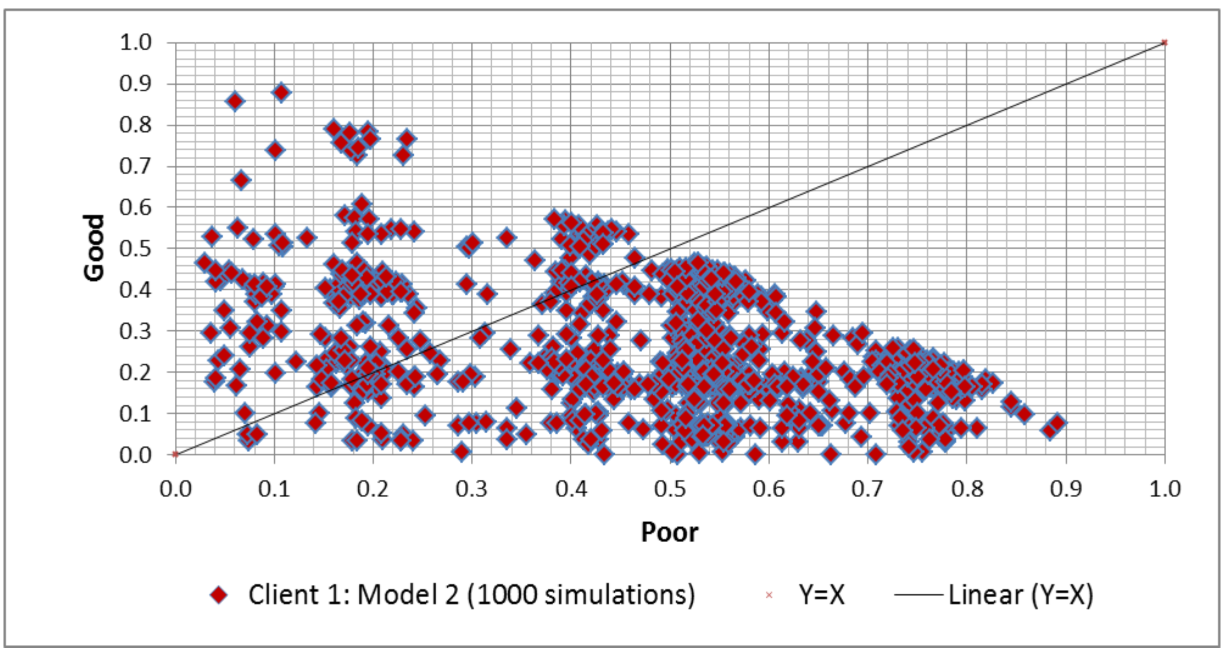

Figure 27 : Plot of weighted number of "Good" tests against "Poor" tests of MODEL 2 for Client 1 (1000 simulations)

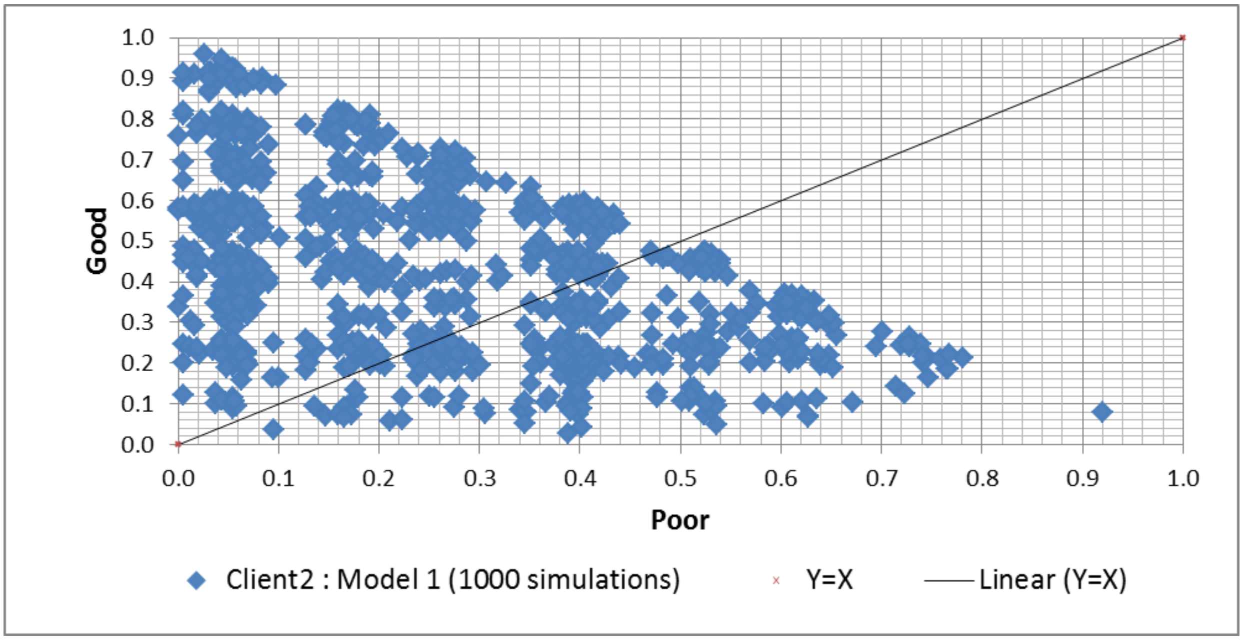

Figure 28 : Plot of weighted number of "Good" tests against "Poor" tests of MODEL 1 for Client 2 (1000 simulations)

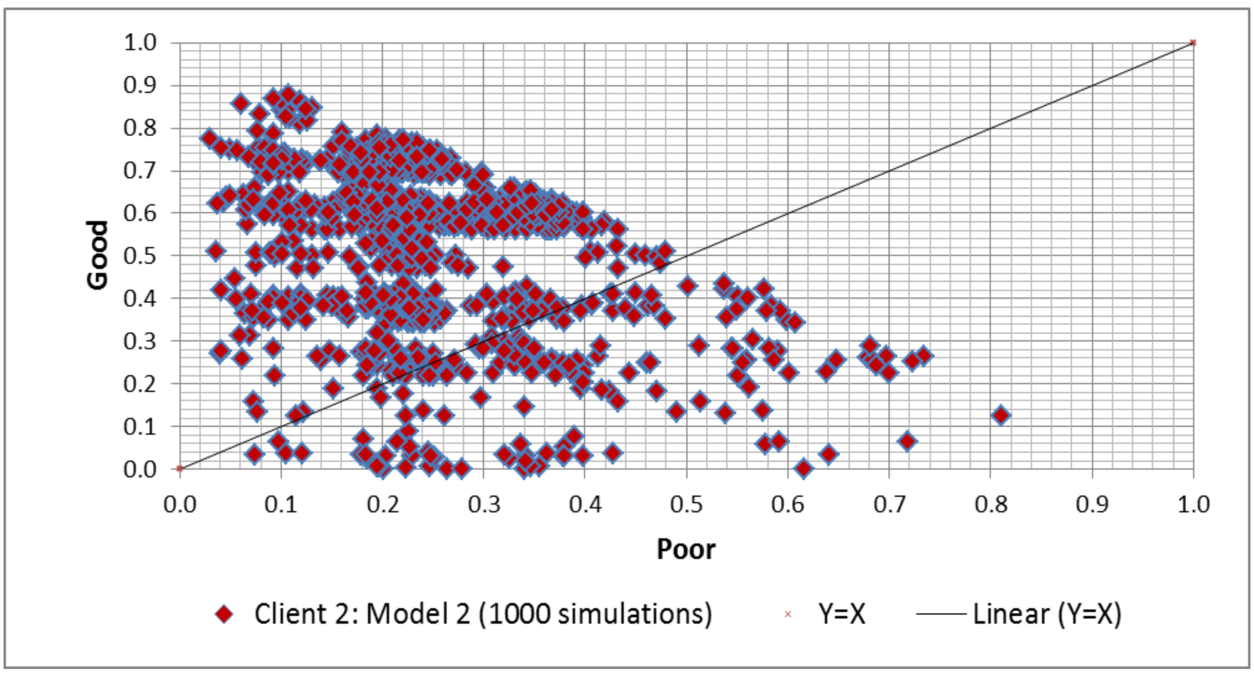

Figure 29 : Plot of weighted number of "Good" tests against "Poor" tests of MODEL 2 for Client 2 (1000 simulations)

Comparing Figure 28 and 29, one may conclude that MODEL 2 performs better than MODEL 1 for Client 2, since MODEL 2 has less data points scattered in the area below the diagonal. However, this contradicts with the observation from the deterministic analysis.

6. CONCLUSION

The previous section has demonstrated the application of two quantitative scoring methods to a real case in practice. One can observe so many interesting findings regarding the suitability of different catastrophe models. These can be processed in practice to particular client and generate relevant conclusions in MSA framework. However, there are some existing limitations and require further research and investigation in the future.

Firstly, the assumptions regarding subjective value within MSA Grid and subjective probability requires more verification in the future. The spread of subjective value is constrained into three deterministic scores. This cannot reflect the level of uncertainty between judgment of “Good”, “Moderate” and “Poor”. For example, some subjective judgments of “Good” may not go to extreme case of score 3 but of score 2.7. In addition, subjective probability can be described in a more complex form which may represent subjective uncertainty more precisely.

Moreover, the quantitative scoring methods in this paper only can be applied on the basis of given MSA Grid, which do not consider whether the MSA Grid given is appropriate. However, the quality of MSA Gird also has great impact on the conclusions derived from scoring catastrophe models. This issue give rise to a wide scope for further research.

Finally, this paper acts as a pioneer product to solve the issue of how to convert qualitative values into quantitative measurements within MSA Grid. Although there are certain scopes for further validation of scoring methods in practice, this paper at least provides simple and approachable way to make decision of suitability of competing catastrophe models at this stage.

7. REFERENCES

1. AIR (2009) Izmit-Düzce Ten Years Later: Is Istanbul at Greater Risk Today? - [Online]. Available: http://www.air-worldwide.com/Publications/AIR-Currents/Izmit-D%CF%8Bzce-Ten-Years-Later--Is-Istanbul-at-Greater-Risk-Today-/ [2009]

2. AIR (2012) About Catastrophe Modeling - [Online]. Available: http://www.air-worldwide.com/Models/About-Catastrophe-Modeling [2012]

3. Berry, Romain (2008), Value-at-Risk: An Overview of Analytical VaR, page 7 , [Online]. Available: http://www.jpmorgan.com/cm/BlobServer/Investment_Analytics_Newsletter_September_2008.pdf?blobkey=id&blobwhere=1158504094293&blobheader=application/pdf&blobheadername1=Cache-Control&blobheadervalue1=private&blobcol=urldata&blobtable=MungoBlobs [2008]

4. CAS (1999) Casualty Actuarial Society -Dynamic Risk Modeling Handbook Chapter 9 [Online]. Available: http://www.casact.org/research/drm/handbook/chapter9.doc[1999]

5. Cooke, R.M. (1991) Experts in Uncertainty, Oxford University Press.

6. Cooke,R.M and Goossens, Louis L.H.J. (2007) , TU Delft expert judgment data base , [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0951832007000944 [15 March 2007]

7. Franco. G. Model Suitability Analysis (MSA) (2012) – own your own view of risk. The Insurance Insider page 11 ,[Online]. Available: http://www.insuranceinsider.com/assets/_files/documents/oct_12/ii__1350896215_Baden-Baden_2012_DAY_1_LR.pdf [21 October 2012]

8. Guy Carpenter (2012), Model Suitability Analysis (MSA), an EMEA Pilot Project for EURO Wind [2012], Unpublished documents

9. Guy Carpenter (2013) , Test C3-3EQ: Frequency and severity catalogue comparison , MSA Standard Templates [2013] , Unpublished documents

10. Li, Xinrong (2013) , SMM522_RPMS Catastrophe Model Suitability Analysis, Student paper [2013], Unpublished documents

11. Major, John A. (2004), Gradients of risk measures: theory and application to catastrophe risk management and reinsurance pricing [Online]. Available: http://www.casact.org/pubs/forum/04wforum/04wf045.pdf [2004]

12. MedCalc, Summary statistics [Online]. Available: http://www.medcalc.org/manual/summary_statistics.php#confidence_interv_SEM#

13. Patricia Grossi, Howard Kunreuther, Chandu C. Patel (2005) Catastrophe Modeling: A New Approach to Managing Risk, Springer Press.

14. Parodi, Pietro (2012) , General Insurance for Pricing, Unpublished notes

15. Peng, Jin (2009) , Value at Risk and Tail Value at Risk in Uncertain Environment [Online]. Available: http://www.orsc.edu.cn/online/090530.pdf [2009]

16. RMS (2008), A Guide to Catastrophe Modeling - [Online]. Available: https://support.rms.com/publications/RMS%20Guide%202008.pdf [2008]

17. RQE (2013), Catastrophe Risk Models & Perils [Online]. Available: http://www.eqecat.com/catastrophe-models [2013]

18. Schwartz, Susan (2012), Costa Rica Seismogenic Zone Structure and Behavior [Online]. Available: http://eps.ucsc.edu/research/focus-on-research-profiles/schwartz-profile.html [2012]