

## RESEARCH INSURANCE ERS

### Authors

Steven Morrison, PhD

Laura Tadrowski, PhD

Moody's Analytics Research

### Contact Us

Americas +1.212.553.1653  
clientservices@moodys.com

Europe +44.20.7772.5454  
clientservices.emea@moodys.com

Asia-Pacific +85.2.2916.1121  
clientservices.asia@moodys.com

Japan +81.3.5408.4100  
clientservices.japan@moodys.com

## Proxy Model Validation

### Overview

In this paper, we discuss the validation of proxy models, commonly used in the insurance industry to replace valuations that would otherwise require Monte Carlo simulation. In practice, proxy model validation inevitably involves a certain amount of subjectivity and is specific to the exact problem at hand. We do not attempt to provide a prescriptive recipe for how validation should be carried out, but rather suggest some general ideas and principles based on our experience implementing proxy models with our clients.

- » It is important to recognise that there is no "one size fits all" approach to validation. Each problem is unique, and validation analysis should be tailored accordingly. Think like a regulator: seek evidence contradicting, rather than supporting, your hypothesis.
- » Use out-of-sample validation, scenarios not used in the fitting process is key. In-sample statistics (such as in-sample  $R^2$ ) do not necessarily provide a good indication of model quality.
- » Think carefully about the scenario budget and its allocation between 'outer' and 'inner' scenarios. Consider the consequences of increasing the budget and shifting allocations – how do your conclusions change?
- » When comparing proxy models with 'actual' (Monte Carlo) valuations recognise and quantify statistical errors in the actual valuations and interpret proxy errors in this context.
- » Ensure the validation stresses provide broad coverage of the risk factor space. Don't just validate where you think the impact is highest, and in particular don't assume the proxy model will identify the correct 'biting scenario'.
- » Materiality of errors should be assessed on an appropriate scale. Think about what the model will be used for and assess errors in this context.
- » Analyse in detail. Understand how the error behaves across the entire risk factor space, and not just on average.

## CONTENTS

1. Introduction .....	3
2. What is out-of-sample validation and why is it important?.....	4
3. Validation scenario budget and allocation.....	8
4. Where to validate?.....	9
5. Performance metrics .....	12
6. Summary.....	15

## 1. Introduction

This paper discusses the validation of proxy models, commonly used in the insurance industry to replace valuations that would otherwise require Monte Carlo simulation. The most common application of such proxy models to date has been the valuation of projected liability and/or asset values in one year's time, in order to calculate 1-year VaR capital and we focus on this application in this paper (though many of the ideas discussed apply more broadly).

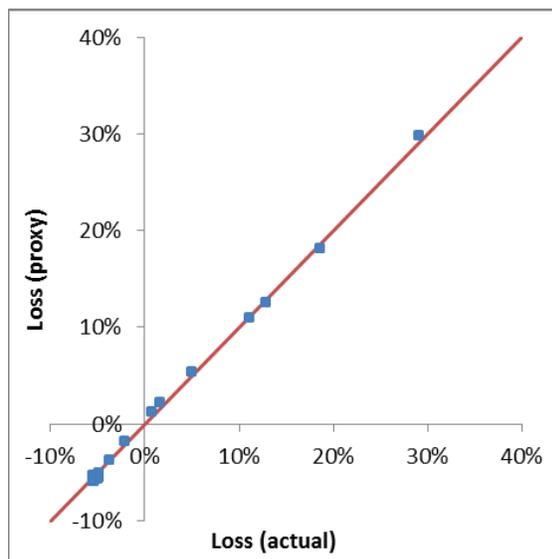
We are fundamentally interested in answering the following question:

For the specific application of interest, is the proxy model suitable for use as an alternative to accurate Monte Carlo valuations<sup>1</sup>?

In this paper we provide guidance on how one can go about answering this question, based on our experience in implementing, calibrating and validating proxy models with a number of firms covering a wide range of insurance business. We do not propose a prescriptive recipe for how the question is answered – in practice, validation inevitably involves a certain amount of subjectivity and is specific to the exact problem at hand. Regulatory guidance is similarly non-prescriptive for this reason. However, we believe that there are some useful general considerations that apply to any proxy validation exercise. In particular, the validation methods and choices described in the note are independent of the specific proxy method used, whether it is Curve Fitting, Least Squares Monte Carlo (LSMC), Replicating Portfolios or any other technique.

The output of a validation exercise is typically a combination of graphs and tables showing how well the proxy model compares to actual Monte Carlo valuations, in a number of out-of-sample 'validation stresses', an example of which is shown in Figure 1.

Figure 1: Example validation analysis



<sup>1</sup> Here, when we refer to 'accurate' Monte Carlo valuations we mean Monte Carlo valuations that have been estimated using a sufficiently large number of risk-neutral scenarios that the resulting statistical errors are negligible.

Stress #	Fund Return	Stressed guarantee cost (as % of initial fund value)			Loss (- $\Delta$ NAV) (as % of initial fund value)		Error (Proxy - Actual):	
		Proxy	Actual	St Error	Proxy	Actual	Relative to initial fund value	Relative to actual loss
1	-39.5%	35.43%	34.63%	0.07%	29.86%	29.06%	0.80%	2.75%
2	-28.5%	23.79%	24.11%	0.05%	18.22%	18.54%	-0.32%	-1.71%
3	-21.8%	18.24%	18.41%	0.06%	12.67%	12.83%	-0.17%	-1.30%
4	-19.6%	16.60%	16.61%	0.08%	11.03%	11.03%	0.00%	-0.03%
5	-10.8%	11.04%	10.55%	0.16%	5.47%	4.97%	0.49%	9.95%
6	-4.2%	7.80%	7.14%	0.21%	2.22%	1.57%	0.66%	41.93%
7	-2.0%	6.87%	6.29%	0.21%	1.30%	0.72%	0.58%	80.57%
8	6.8%	3.88%	3.45%	0.19%	-1.69%	-2.13%	0.43%	-20.36%
9	15.7%	1.86%	1.78%	0.14%	-3.71%	-3.80%	0.08%	-2.17%
10	24.5%	0.62%	0.66%	0.09%	-4.96%	-4.91%	-0.04%	0.85%
11	31.1%	0.09%	0.69%	0.09%	-5.49%	-4.88%	-0.61%	12.42%
12	33.3%	-0.03%	0.53%	0.07%	-5.60%	-5.05%	-0.55%	10.94%
13	42.1%	-0.24%	0.18%	0.04%	-5.81%	-5.39%	-0.42%	7.83%
14	50.9%	-0.18%	0.08%	0.03%	-5.75%	-5.49%	-0.26%	4.69%
15	59.8%	0.02%	0.03%	0.01%	-5.55%	-5.55%	-0.01%	0.15%
16	66.4%	0.18%	0.01%	0.01%	-5.40%	-5.56%	0.17%	-2.99%
17	68.6%	0.22%	0.01%	0.01%	-5.35%	-5.56%	0.21%	-3.73%
18	77.4%	0.30%	0.03%	0.02%	-5.27%	-5.55%	0.28%	-4.98%
19	86.2%	0.17%	0.00%	0.00%	-5.40%	-5.57%	0.17%	-3.06%
20	95.1%	-0.27%	0.00%	0.00%	-5.85%	-5.57%	-0.27%	4.93%
Average							0.06%	6.83%
Average absolute							0.33%	10.87%
Maximum absolute							0.80%	80.57%

The paper will describe some of the main considerations involved in producing and interpreting this information, and is organized as follows:

- » **Section 2** motivates and describes the principle of out-of-sample validation.
- » **Section 3** considers the choice of how many stresses and scenarios to use in the validation.
- » **Section 4** considers the question of where to validate i.e. the selection of validation stresses.
- » **Section 5** describes the measurement of model performance.
- » **Section 6** concludes.

## 2. What is out-of-sample validation and why is it important?

In this paper, we focus on out-of-sample (OOS) validation, here defined as the process of comparing proxy model valuations with simulation-based valuations calculated using risk-neutral scenarios that were not used in the fitting process. In our experience, OOS validation has become the standard approach to proxy model validation, in particular as part of regulatory Internal Model Approval processes<sup>2</sup>. In this section, we demonstrate why in-sample diagnostics are insufficient to draw any meaningful conclusions regarding proxy model performance, and motivate the use of OOS validation.

As set out in the introduction, the methods described in this note are independent of choice of proxy method. However, in order to present some practical examples we will consider and compare two specific proxy methods here: Curve Fitting and Least Squares Monte Carlo (LSMC).

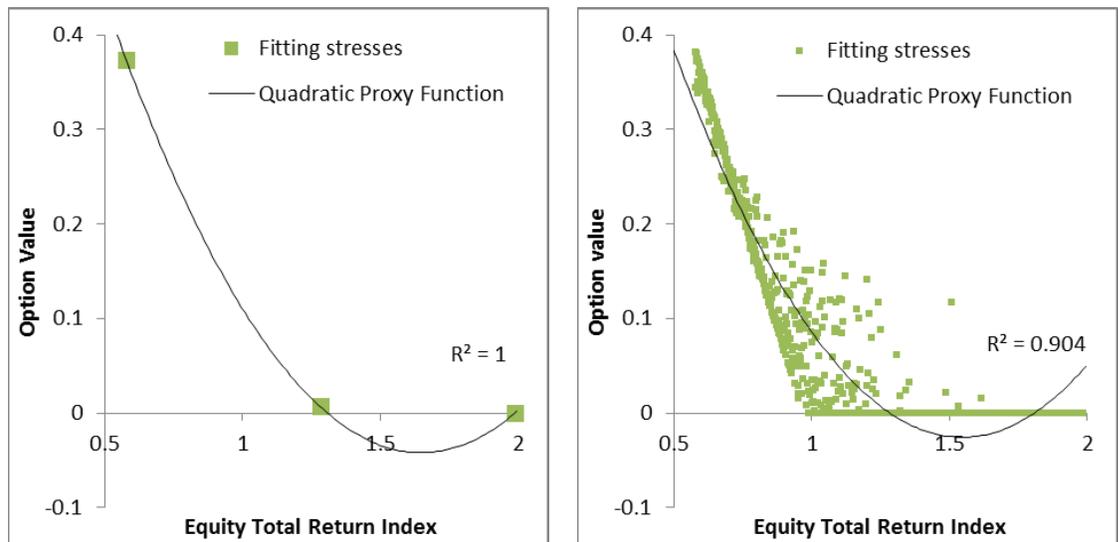
<sup>2</sup> See for example "Solvency II: event on technical issues for life insurers intending to use an internal model", Prudential Regulation Authority.

The Curve Fitting and LSMC methods can be considered as different ways of implementing the same basic proxy method. In both cases, a function is fitted to valuation data produced using a nested simulation exercise. In order to be able to run the nested simulation in available timescales, a compromise is made either on the number of outer scenarios (in this context sometimes called “fitting stresses”) or the number of risk-neutral inner scenarios (in this context sometimes called “fitting scenarios”). The term Curve Fitting is usually used to describe the first type of compromise: we fit a proxy function exactly in a small number of fitting stresses that have been very accurately valued using the cash flow model. This corresponds to a nested simulation exercise with a low number of outer scenarios and a high number of inner scenarios. At the other extreme, the term Least Squares Monte Carlo is used to describe fitting a proxy function (through least squares regression or other similar techniques) through a very large number of relatively inaccurate valuations. This corresponds to a nested simulation exercise with a high number of outer scenarios and a low number of inner scenarios.

As an illustrative example, polynomial proxy functions are fitted to the value of a 1-year put option on an equity index, with a strike of 1 (initially at-the-money). We use a Black-Scholes valuation model with a fixed interest-rate and implied volatility, so that the put option value depends on a single risk factor: the underlying equity total return index.

Figure 2 shows two different proxy function fits, produced using the Curve Fitting method (left-hand chart) and LSMC method (right-hand chart)<sup>3</sup>. In both cases a quadratic (order 2) polynomial has been fit. The fitted quadratic functions are different using the two different methods. From this information alone, can we infer whether either of these proxy functions are fit for purpose?

Figure 2: Quadratic function fits: Curve Fitting method (left-hand chart) and LSMC method (right-hand chart)



It may be tempting to trust the fit produced by the Curve Fitting method more than that produced by the LSMC method. After all, we know that this function exactly fits the ‘actual’ value at a number of fitting stresses, whereas we don’t have any ‘actual’ values to compare the LSMC proxy function to (as these valuations have been produced using only 2 inner scenarios per fitting stress, and so are subject to relatively large statistical error)<sup>4</sup>. However, we only know that the Curve Fitting method produces an exact fit at three

<sup>3</sup> The curve fitting method here uses 3 fitting stresses, with 1,000 inner scenarios per stress, while the LSMC method uses 1,000 fitting stresses with 2 scenarios per stress. For both methods, inner scenarios are sampled in antithetic pairs.

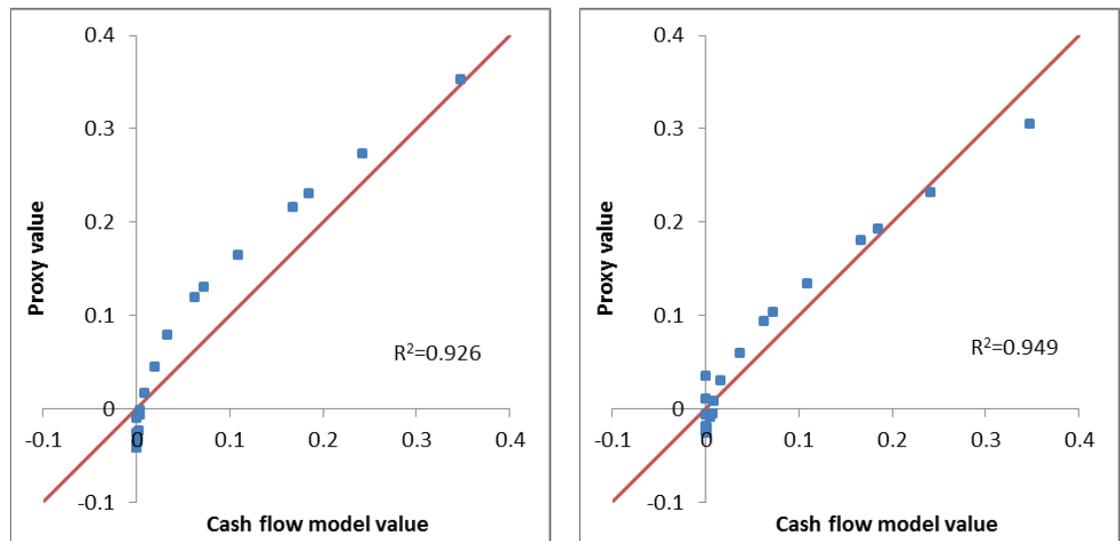
<sup>4</sup> Actually, the valuations used to fit the proxy function in the Curve Fitting method are also estimated using simulation, albeit with a relatively large number of inner scenarios (1,000 per stress in the examples here). So even in the Curve Fitting case we don’t have any completely accurate ‘actual’ values to compare the proxy to, though (assuming enough inner scenarios have been used) we should have very accurate approximations.

stresses (in this case, chosen to be equity index values of 0.58, 1.28 and 1.99), while we really want to apply the fitted proxy function over a much larger number of stresses. Of course, we can't compare the proxy function values to the cash flow model values in all stresses that may be of interest (as this would involve a full nested stochastic simulation, with a large number of both outer and inner scenarios), but we can validate in a small number of representative validation stresses. By using the same validation stresses to evaluate both methods, and using scenarios not used in the fitting process (i.e. out-of-sample data), we can directly assess the performance of both methods on a like-for-like basis.

An example of such an out-of-sample-validation is shown in Figure 3, which plots proxy against cash flow model values in 20 validation stresses, chosen to uniformly span the range of equity index values used in the fitting process. The cash flow model value in each validation stress is estimated using 1,000 risk-neutral inner scenarios.

Given these results, we conclude that neither the Curve Fitting nor LSMC method has performed particularly well, and it is unlikely that either would satisfy regulatory validation requirements. While we would caution against using any single summary statistic to quantify the validation, it is interesting here to look at the out-of-sample  $R^2$  (measured across the 20 validation stresses) and compare this to the in-sample  $R^2$  (measured across all of the fitting stresses). In this example, the curve fitting method has an out-of-sample  $R^2$  of 0.926 (compared to an in-sample  $R^2$  of 1), while the LSMC method has an out-of-sample  $R^2$  of 0.949 (compared to an in-sample  $R^2$  of 0.904). This illustrates a general feature of these particular types of proxy methods, with the curve fitting method never producing a higher out-of-sample  $R^2$  than its in-sample  $R^2$  (which is always 1 by construction). In contrast, the LSMC method may result in a far higher out-of-sample  $R^2$  than its in-sample  $R^2$ , simply because the in-sample statistic is calculated using a lower number of inner scenarios (and hence noisier cash flow model estimates).

Figure 3: Validation of quadratic proxy function fits: Curve Fitting method (left-hand chart) and LSMC method (right-hand chart)



Since the choice of quadratic polynomial doesn't appear to capture the true shape of the option value particularly well (over the range of validation stresses considered here), we consider raising the order of the polynomial in an attempt to improve the quality of fit. Figure 4 shows the corresponding fits using order 4 (quartic) polynomials, with an additional two fitting stresses being used for the curve fit, and the same 1,000 fitting stresses as before being used for the LSMC fit. The corresponding out-of-sample-validations are shown in Figure 5, and indicate that both quartic polynomials provide a far improved fit to the actual values

(compared to the quadratic fits) in a wide range of validation stresses. Out-of-sample  $R^2$  are 0.998 (Curve Fitting) and 0.999 (LSMC) compared to in-sample  $R^2$  of 1 (Curve Fitting) and 0.951 (LSMC). As before, we note that the in-sample  $R^2$  are of limited use in assessing quality of fit, always overstating fitting quality in the case of curve fitting, and often significantly understating fitting quality in the case of LSMC. The only way to truly assess fitting quality is via out-of-sample validation, which in this case concludes that both methods produce similarly accurate proxy functions, despite the differences in the in-sample data used.

Figure 4: Quartic proxy function fits: Curve Fitting method (left-hand chart) and LSMC method (right-hand chart)

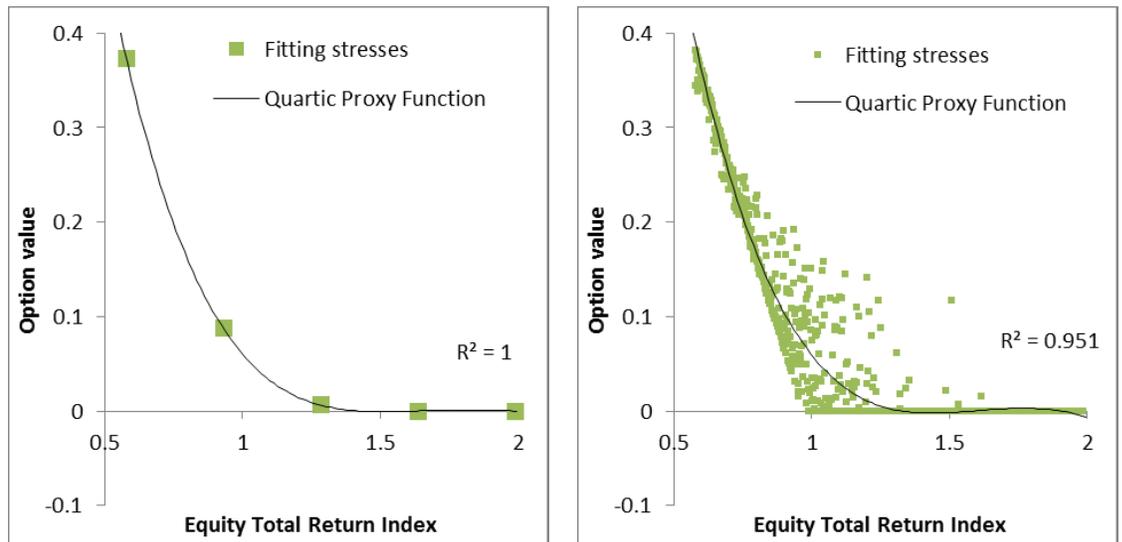
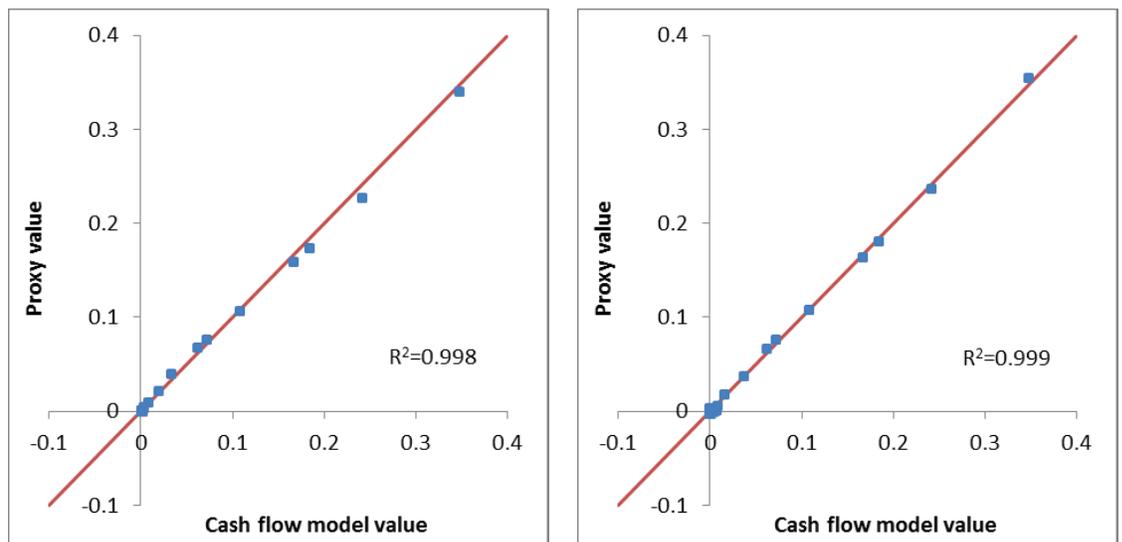


Figure 5: Validation of quartic proxy function fits: Curve Fitting method (left-hand chart) and LSMC method (right-hand chart)



While the above discussion has focused on proxy methods of the Curve Fitting and Least Squares Monte Carlo type, similar conclusions are likely to hold for any proxy method. In particular, the in-sample  $R^2$  measures commonly used to assess quality of fit using the Replicating Portfolio method are similarly flawed as an indicator of out-of-sample performance.<sup>5</sup>

<sup>5</sup> See Barrie & Hibbert Model Insights, "Replicating Portfolios for Economic Capital: Replication or Approximation?" (October 2008).

### 3. Validation scenario budget and allocation

There are a number of choices that the proxy function user needs to consider when implementing an out-of-sample validation process. Here we consider one of the first choices the user is likely to make: The choice of how many out-of-sample validation scenarios to use, and also how to allocate these between 'outer' and 'inner' scenarios.

In the previous section, we described the Curve Fitting and LSMC fitting processes in terms of nested stochastic simulation, with different compromises made in the number of inner and outer scenarios. Similarly, the general out-of-sample validation process can be thought of as a nested stochastic simulation: OOS validation involves carrying out accurate valuations (using a relatively large number of risk-neutral inner scenarios) under a number of representative stresses to the relevant risk factors (outer scenarios).

The total number of scenarios involved in an OOS validation exercise is:

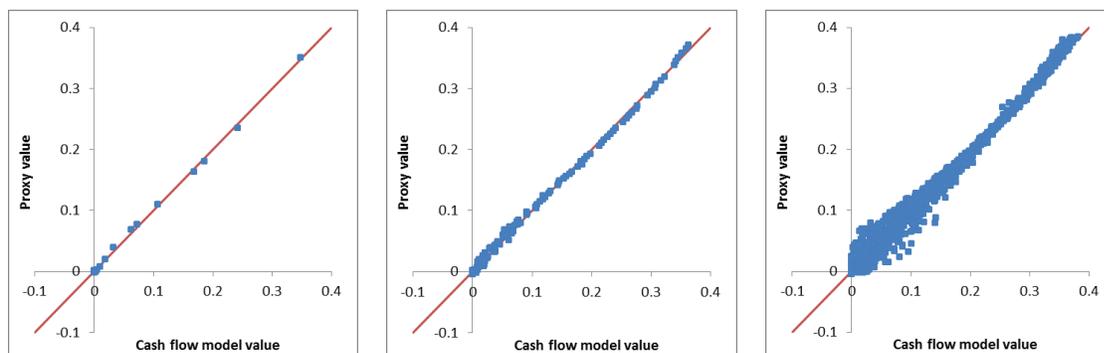
$$\begin{aligned} \text{Total number of validation scenarios} \\ &= \text{Number of validation stresses ('outers')} \\ &\times \text{Number of validation scenarios per stress ('inners')} \end{aligned}$$

The total number of scenarios is naturally constrained by available computational resources, and in practice this means that firms will often run as many as possible within some fixed timescale. However, arguing insufficient resources to perform adequate testing of proxy model performance is unlikely to satisfy regulatory requirements. A minimal total number of scenarios will be required in order to draw meaningful conclusions about quality of fit (and hence gain model approval).

Given a fixed computational budget (and therefore total scenario budget), how do we optimally allocate this between outers (stresses) and inners? In the example validations shown in Figure 3 and Figure 5, we chose a relatively large number of inner scenarios (1,000) to perform accurate valuations under 20 stresses, involving  $20 \times 1,000 = 20,000$  scenarios in total. We now ask the question: assuming that our scenario budget is 20,000, how does the resulting validation information change as we shift the outer/inner scenario mix?

Figure 6 shows three different validation charts for the same proxy function<sup>6</sup>. Each validation is produced using the same total validation scenario budget (20,000) but under different allocations (20, 200 and 2,000 outers)<sup>7</sup>.

Figure 6: Example validation charts for three different scenario allocations: 20 x 1,000 (left); 200 x 100 (center); 2,000 x 10 (right)



<sup>6</sup> The proxy function here is an order 4 polynomial fitted using the LSMC method, representing the value of an at-the-money put option.

<sup>7</sup> Here and in most other charts of this type shown in this paper, we have not plotted statistical confidence intervals around the cash flow model values (because they would be hard to read and/or difficult to estimate), but it should be remembered that these values are just estimates and subject to statistical error (particularly on the right-most chart, where only 10 risk-neutral scenarios have been used).

At one extreme (left-most chart), the fit looks good in each stress. However, in this case we only know how well the model works at a relatively small number of stresses. We can try and counter this by increasing the number of stresses, while decreasing the number of valuation scenarios (inners) to compensate. More points appear on the scatter chart, giving a wider coverage of the risk driver space but the noise on the cash flow model values increases as we do this. The right-most chart tests the quality of fit in a very large number of stresses, but the quality of fit in any individual stress is hard to assess (since only 10 risk-neutral inner scenarios have been used to estimate the cash flow model values). There are some seemingly large errors though these may or may not be due to noise.

We present the question: Which (if any) of the charts in Figure 6 would lead to the conclusion that the proxy model performs well? Which chart provides the most useful validation information?

In our experience, practitioners will often already have a view regarding the number of inner scenarios that they typically require to 'accurately' value a particular product or book of business (perhaps based on previous valuation exercises carried out in some other context). However, in the context of proxy model validation, it is worth assessing the validity of this assumption and putting a quantitative framework around accuracy requirements. In particular, proxy model users should ask whether it is possible to decrease the number of inner scenarios (thereby increasing the number of validation stresses and thus gaining further confidence in the range of stresses over which the proxy function 'works'). On the other hand, it is important to ensure that the number of inner scenarios is sufficiently large such that one can draw meaningful conclusions about the size of any proxy errors relative to Monte Carlo noise. Regardless of how many inner scenarios are chosen it is important to try and understand the amount of statistical uncertainty in 'actual' values, and to assess the quality of any proxy errors in this context. We will come back to this point in Section 5.

#### 4. Where to validate?

Having chosen the number of validation stresses, where should we place these in the risk factor space?

Ultimately, we want to test that the function is sufficiently accurate in all regions of the risk driver space where it may be applied. Given constraints on scenario budget, it would be inefficient to place stresses at points where the function is not required (for example points of extremely low probability and/or low impact). On the other hand, it may also be inefficient to place too many points too closely in one region of the risk factor space (leaving other regions empty) as this doesn't provide complete coverage of all possible states of the world.

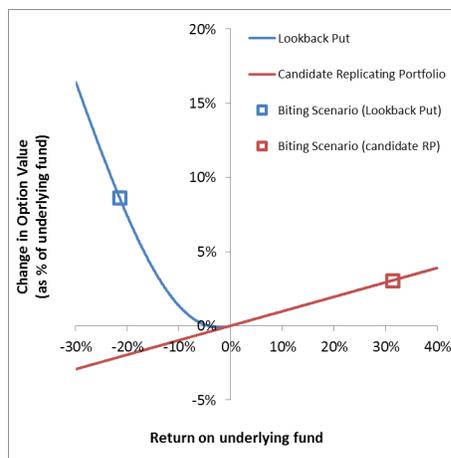
#### **A RISKY STRATEGY: ONLY VALIDATING WHERE THE PROXY FUNCTION INDICATES HIGH IMPACT**

A particularly high risk strategy would be to use the un-validated proxy function to identify a region of the space believed to be of particularly high impact (e.g. scenarios that the proxy function predicts will give rise to losses at the 99.5<sup>th</sup> percentile, or thereabouts) and then to validate the proxy model only in that region. If the proxy function performs well in this region this doesn't necessarily imply that it identified the 'right' region in the first place, and there is a risk that the proxy function performs well in a region that isn't interesting for the particular application that it is being used for.

An illustrative example demonstrating the potential drawbacks of this approach is presented in Figure 7. We consider valuation of a lookback put option maturing in 2 years<sup>8</sup>. The lookback option has a similar payoff to a vanilla put option. However, the 'strike' of this option is not fixed but set at the maximum value of the underlying fund achieved over the life of the option. This 'ratchet' feature is common in many guarantees embedded in insurance products, in particular variable annuities.

The profile of the change in option value, under stresses to the underlying fund, is shown in Figure 7. Like vanilla put options, the cost rises as the underlying fund falls (and the intrinsic value increases); however, the lookback option is also exposed to *rising* fund values (as the time value increases due to the increasing strike). Focusing on the region of positive fund returns, the change in value of the lookback option is actually a linear function of the fund return so can be *perfectly replicated* by a position in the underlying fund (or, equivalently, perfectly proxied by a linear function).

Figure 7: Profile of lookback put option and replicating portfolio in underlying fund



Suppose we use this candidate Replicating Portfolio as a proxy for the lookback option. This replicating portfolio is perfect as long as the fund rises, but an increasingly poor proxy for increasingly large falls in the fund (underestimating the change in option value in such cases). If we use this proxy model to identify the 99.5% increase in guarantee cost, this is identified as the scenario where the fund *rises* by 31%, with a corresponding increase in guarantee cost of 3% (of the initial fund value). The proxy model would validate perfectly in this particular 'biting scenario'.

However, although this is the biting scenario for the RP, it isn't the biting scenario for the lookback option. The actual 99.5% increase in guarantee cost corresponds to a 22% *fall* in the fund value, with a corresponding increase in guarantee cost of 9%. In summary, the linear proxy model would lead to the conclusion that the greatest losses are associated with the highest returns on the underlying fund; when in fact the greatest losses are actually driven by very large negative returns.

<sup>8</sup> Here the underlying fund is assumed to follow a Black-Scholes model with a volatility of 10% and risk-free rate of 2%.

## GUIDELINES FOR SELECTION OF VALIDATION STRESSES

Though the above example may seem contrived, it illustrates in an extreme way how restricting validation to a region of space where the proxy function suggests the impact is largest is a dangerous validation strategy. Validation may be perfectly adequate (or indeed perfect) in certain regions but extremely poor in others, and putting too much faith in intuition as to where impact is large exposes the modeler to the risk that their intuition is flawed. Although it can be useful to hand-pick some validation stresses in regions where impact is *expected* to be large, in our view validation needs to test the quality of fit in a much wider range of possible outcomes.

In our experience, good validation practice involves selecting stresses which are a combination of two broad types:

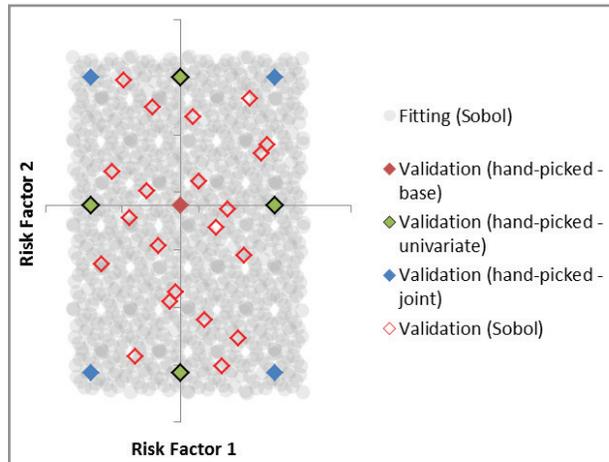
- 1 **'Hand-picked' stresses** i.e. stresses in regions that are expected to be of particular relevance, motivated either using a model, or on intuition based on understanding of the particular product/business features. In a VaR application, these often represent 'extreme' stresses, for example 99.5% increases and falls in the risk factors (the above example highlights the importance of considering extreme movements in both directions). Hand-picked stresses may be univariate i.e. stress individual risk factors with the others held fixed and/or joint stresses i.e. stresses to a number of risk factors simultaneously. A 'base' stress (or non-stress), i.e. valuation under no change to risk factors, is often included here.
- 2 **Random (or quasi-random) joint stresses.** These may be generated by an Economic Scenario Generator (for example the real-world ESG that will be used to project the proxy function), but more commonly are sampled uniformly within the range of risk factor stresses used in fitting the proxy function. The aim of the random stresses is to fill the risk factor space in such a way that we can be confident in wide applicability of the proxy function (and not just in the extremes for example).

Note that in practice the volume of the risk factor space we are seeking to fill increases rapidly with the number of risk factors. However, the number of validation stresses that we have seen firms typically run (given typical scenario budgets) have been around 50, often resulting in large 'gaps' in the risk-factor space (where the dimension could be 10 or more). In such cases, there is a risk that the validation overstates the quality of fit simply due to the particular choice of validation stresses used, and another set of validation stresses (chosen using similar principles) would result in a far poorer validation. Sobol sequences (rather than pseudo-random numbers) are often used to sample from high dimension multivariate uniform distributions in a relatively efficient way (intuitively, by minimizing the size of gaps in the space).

An example choice of validation stresses, covering both types described above, is shown in Figure 8 (in two risk factor dimensions). The stresses used in the fitting process are also shown for comparison (in this case 1,000 fitting stresses sampled from multivariate uniform distribution using Sobol). Note that the hand-picked scenarios here are 'extreme' in the sense that they are close to the edges/corners of the range of fitting stresses. However, they do lie within the range of fitting stresses. In general, we would caution against choosing validation scenarios that lie outwith the range of stresses used in the fitting process (or on the boundary). If the model user wants to hand pick a validation stress outwith the fitting range, this is perhaps an indication that the range of fitting stresses is inappropriate, in which case this range should be expanded to cover the chosen set of validation stresses, and the proxy model re-fit<sup>9</sup>.

<sup>9</sup> Although the process of fitting needs to be carried out before the proxy model can be validated, the modeler should really consider selection of validation stresses at the same time as selecting the fitting stresses. In an ideal implementation, the choice of validation stresses (or more generally, consideration of where the proxy function will be applied) should guide the choice of fitting stresses (rather than choosing the validation stresses after the fit has been carried out).

Figure 8: Example choice of validation stresses in 2 risk factors



## 5. Performance metrics

Having chosen validation stresses, we can proceed to calculate 'actual' (i.e. accurate) values in those stresses and compare with values predicted by the proxy function. We have already considered a visual comparison of values using scatter plots (see Figure 5 for example) but in practice this is usually accompanied by a more detailed analysis in the form of a table quoting values in each valuation stress along with some error metrics. There are a number of considerations in making this comparison, in particular:

- 1 Which values should be compared? Often the proxy model itself describes the market-consistent value of liabilities only (and sometimes a separate proxy function describes the market value of assets), but the ultimate quantity of interest may be a combination of these. For example, in a 1-year VaR capital calculation (such as the Solvency 2 SCR), the loss is usually defined as minus the *change in net asset value* (relative to base) and therefore this variable may be the most useful to focus attention on.
- 2 How is the proxy value compared to the actual value? Should we focus attention on the error (i.e. monetary difference in values), or the 'relative error' (i.e. difference relative to something else)? If the latter, what should we measure the error relative to?

Figure 9 shows an example validation table, corresponding to the put option example considered in Section 2. The proxy function here is a quartic polynomial in a single risk factor (the underlying fund), fitted using LSMC, and we compare in 20 stresses chosen to uniformly span the range of fitting scenarios. Stresses in Figure 9 are ordered in terms of the fund return (the single risk factor in this example).

Figure 9: Example proxy validation table (one year put option example; quartic proxy function)

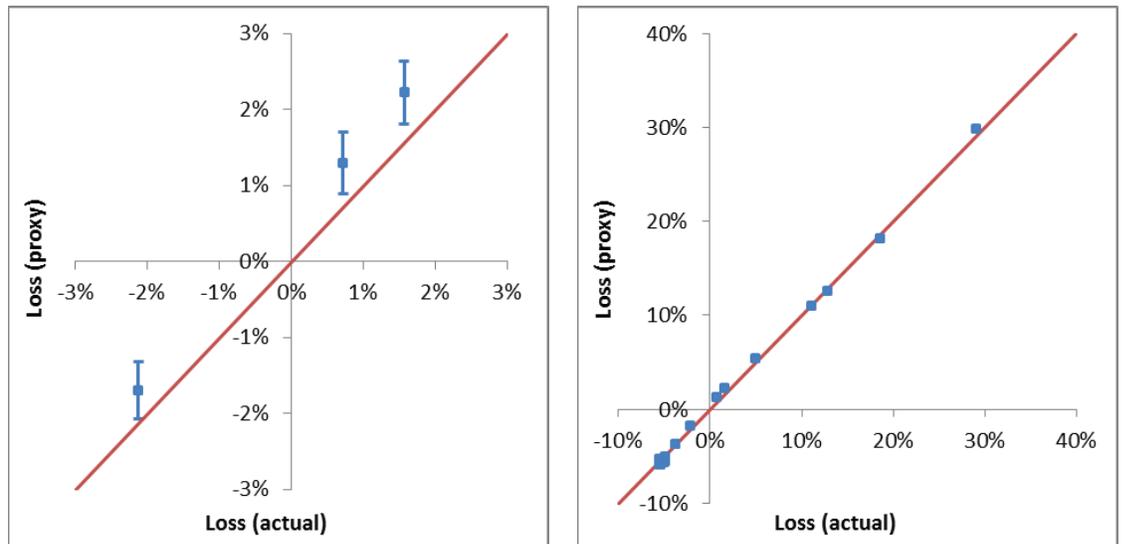
Stress #	Fund Return	Stressed guarantee cost (as % of initial fund value)			Loss (- $\Delta$ NAV) (as % of initial fund value)		Error (Proxy - Actual):	
		Proxy	Actual	St Error	Proxy	Actual	Relative to initial fund value	Relative to actual loss
1	-39.5%	35.43%	34.63%	0.07%	29.86%	29.06%	0.80%	2.75%
2	-28.5%	23.79%	24.11%	0.05%	18.22%	18.54%	-0.32%	-1.71%
3	-21.8%	18.24%	18.41%	0.06%	12.67%	12.83%	-0.17%	-1.30%
4	-19.6%	16.60%	16.61%	0.08%	11.03%	11.03%	0.00%	-0.03%
5	-10.8%	11.04%	10.55%	0.16%	5.47%	4.97%	0.49%	9.95%
6	-4.2%	7.80%	7.14%	0.21%	2.22%	1.57%	0.66%	41.93%
7	-2.0%	6.87%	6.29%	0.21%	1.30%	0.72%	0.58%	80.57%
8	6.8%	3.88%	3.45%	0.19%	-1.69%	-2.13%	0.43%	-20.36%
9	15.7%	1.86%	1.78%	0.14%	-3.71%	-3.80%	0.08%	-2.17%
10	24.5%	0.62%	0.66%	0.09%	-4.96%	-4.91%	-0.04%	0.85%
11	31.1%	0.09%	0.69%	0.09%	-5.49%	-4.88%	-0.61%	12.42%
12	33.3%	-0.03%	0.53%	0.07%	-5.60%	-5.05%	-0.55%	10.94%
13	42.1%	-0.24%	0.18%	0.04%	-5.81%	-5.39%	-0.42%	7.83%
14	50.9%	-0.18%	0.08%	0.03%	-5.75%	-5.49%	-0.26%	4.69%
15	59.8%	0.02%	0.03%	0.01%	-5.55%	-5.55%	-0.01%	0.15%
16	66.4%	0.18%	0.01%	0.01%	-5.40%	-5.56%	0.17%	-2.99%
17	68.6%	0.22%	0.01%	0.01%	-5.35%	-5.56%	0.21%	-3.73%
18	77.4%	0.30%	0.03%	0.02%	-5.27%	-5.55%	0.28%	-4.98%
19	86.2%	0.17%	0.00%	0.00%	-5.40%	-5.57%	0.17%	-3.06%
20	95.1%	-0.27%	0.00%	0.00%	-5.85%	-5.57%	-0.27%	4.93%
Average							0.06%	6.83%
Average absolute							0.33%	10.87%
Maximum absolute							0.80%	80.57%

In this table, we quote two types of value: (1) the 'stressed guarantee cost', i.e. market-consistent value of the put option under stress (highlighted in green) and (2) the 'loss' i.e. the change in guarantee cost relative to base (highlighted in orange). Although the stressed guarantee cost is the variable that is directly described by the proxy function, as discussed above we find it useful to subtract off the base guarantee cost (5.57% of the initial fund value) as it is this *change* in guarantee cost that defines the VaR capital requirement.

We have also quoted two measures of proxy model error: (1) the error relative to the initial (i.e. unstressed) fund value, and (2) the error relative to the actual loss in each scenario. The first error can be scaled into a monetary error simply by scaling up by the monetary value of the fund. The second of these error measures may intuitively seem useful, and is often quoted, but should be interpreted with care. The seemingly large relative errors in stresses 6-8 are partly a reflection of the very small loss in these stresses (as they correspond to very small stresses to the equity fund). Focusing on proxy vs actual losses in these stresses (left-hand chart in Figure 10) we see that the errors here do indeed appear large in relative terms. However, on a scale that includes losses in all validation stresses these errors appear small (right-hand chart in Figure 10)<sup>10</sup>. Since the VaR capital requirement corresponds to a relatively large stress, it is arguably more relevant to assess materiality of errors on this larger scale. For similar reasons, some firms also find it useful to express errors relative to other key balance sheet items, such as the base (unstressed) market-consistent value of liabilities, or the base guarantee cost.

<sup>10</sup> Note that estimated 95% confidence intervals on 'actual' losses are shown on the left-hand chart but omitted from the right-hand chart as they are small on this scale.

Figure 10: Proxy vs actual loss: Validation stresses 6-8 (left-hand chart); All validation stresses (right-hand chart)



As noted in Section 3, the 'actual' numbers here are all estimated using a finite (albeit large) number of risk-neutral scenarios and so subject to sampling error. Note that many of the errors here are statistically significant. For example, in stress 1 (the stress with the largest fall in fund value, and hence the largest increase in guarantee cost), the proxy error is around 12 times the estimated standard error (note that standard errors are shown in the 5th column of Figure 9). However, statistically significant error isn't an issue in itself – after all, no proxy function is likely to be perfect and any (non-zero) proxy error can be made increasingly large relative to the standard error simply by running more inner scenarios. The important question here is whether the firm assess proxy errors to be acceptably small, and (as discussed in Section 3), it is important that standard errors are sufficiently small for firms to be able to make this assessment. The basic principle to bear in mind here is that the assessment of whether or not the proxy function is successful should not depend on the random number seeds used in the estimation of the 'actual' valuations.

In addition to measuring errors in each individual stress, it can be useful to summarize this information using metrics such as the average and maximum absolute error, though we recommend that an analysis of errors in individual stresses should also be undertaken as part of any validation exercise. In particular, it is useful to investigate whether there are any particular areas of the risk factor space where errors are systematically large, as this may indicate systematic issues in the fitting process and suggest areas for improvement (for example in selection of location of fitting stresses).

## 6. Summary

In this paper, we have discussed the validation of proxy models, commonly used in the insurance industry to replace valuations that would otherwise require Monte Carlo simulation. In practice, proxy model validation inevitably involves a certain amount of subjectivity and is specific to the exact problem at hand. As such, we have not attempted to provide a prescriptive recipe for how validation should be carried out, but rather suggest some general ideas and principles based on our experience implementing proxy models with our clients.

Our key messages are summarized below:

- » Firstly, recognise that there is no "one size fits all" approach to validation. Each problem is unique, and validation analysis should be tailored accordingly. Think like a regulator: seek evidence contradicting, rather than supporting, your hypothesis.
- » Out-of-sample validation (i.e. use of scenarios not used in the fitting process) is key. In-sample statistics (such as in-sample  $R^2$ ) do not necessarily provide a good indication of model quality.
- » Think carefully about scenario budget and its allocation between 'outer' and 'inner' scenarios. Consider the consequences of increasing budget and shifting allocations – how do your conclusions change?
- » When comparing proxy models with 'actual' (Monte Carlo) valuations recognise and quantify statistical errors in the actual valuations and interpret proxy errors in this context.
- » Ensure validation stresses provide broad coverage of the risk factor space. Don't just validate where you think the impact is highest, and in particular don't assume the proxy model will identify the correct 'biting scenario'.
- » Materiality of errors should be assessed on an appropriate scale. Think about what the model will be used for and assess errors in this context.
- » Analyse in detail. Understand how the error behaves across the entire risk factor space, and not just on average.

© 2015 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "MOODY'S"). All rights reserved.

CREDIT RATINGS ISSUED BY MOODY'S INVESTORS SERVICE, INC. ("MIS") AND ITS AFFILIATES ARE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES, AND CREDIT RATINGS AND RESEARCH PUBLICATIONS PUBLISHED BY MOODY'S ("MOODY'S PUBLICATIONS") MAY INCLUDE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES. MOODY'S DEFINES CREDIT RISK AS THE RISK THAT AN ENTITY MAY NOT MEET ITS CONTRACTUAL, FINANCIAL OBLIGATIONS AS THEY COME DUE AND ANY ESTIMATED FINANCIAL LOSS IN THE EVENT OF DEFAULT. CREDIT RATINGS DO NOT ADDRESS ANY OTHER RISK, INCLUDING BUT NOT LIMITED TO: LIQUIDITY RISK, MARKET VALUE RISK, OR PRICE VOLATILITY. CREDIT RATINGS AND MOODY'S OPINIONS INCLUDED IN MOODY'S PUBLICATIONS ARE NOT STATEMENTS OF CURRENT OR HISTORICAL FACT. MOODY'S PUBLICATIONS MAY ALSO INCLUDE QUANTITATIVE MODEL-BASED ESTIMATES OF CREDIT RISK AND RELATED OPINIONS OR COMMENTARY PUBLISHED BY MOODY'S ANALYTICS, INC. CREDIT RATINGS AND MOODY'S PUBLICATIONS DO NOT CONSTITUTE OR PROVIDE INVESTMENT OR FINANCIAL ADVICE, AND CREDIT RATINGS AND MOODY'S PUBLICATIONS ARE NOT AND DO NOT PROVIDE RECOMMENDATIONS TO PURCHASE, SELL, OR HOLD PARTICULAR SECURITIES. NEITHER CREDIT RATINGS NOR MOODY'S PUBLICATIONS COMMENT ON THE SUITABILITY OF AN INVESTMENT FOR ANY PARTICULAR INVESTOR. MOODY'S ISSUES ITS CREDIT RATINGS AND PUBLISHES MOODY'S PUBLICATIONS WITH THE EXPECTATION AND UNDERSTANDING THAT EACH INVESTOR WILL, WITH DUE CARE, MAKE ITS OWN STUDY AND EVALUATION OF EACH SECURITY THAT IS UNDER CONSIDERATION FOR PURCHASE, HOLDING, OR SALE.

MOODY'S CREDIT RATINGS AND MOODY'S PUBLICATIONS ARE NOT INTENDED FOR USE BY RETAIL INVESTORS AND IT WOULD BE RECKLESS FOR RETAIL INVESTORS TO CONSIDER MOODY'S CREDIT RATINGS OR MOODY'S PUBLICATIONS IN MAKING ANY INVESTMENT DECISION. IF IN DOUBT YOU SHOULD CONTACT YOUR FINANCIAL OR OTHER PROFESSIONAL ADVISER.

ALL INFORMATION CONTAINED HEREIN IS PROTECTED BY LAW, INCLUDING BUT NOT LIMITED TO, COPYRIGHT LAW, AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT.

All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, all information contained herein is provided "AS IS" without warranty of any kind. MOODY'S adopts all necessary measures so that the information it uses in assigning a credit rating is of sufficient quality and from sources MOODY'S considers to be reliable including, when appropriate, independent third-party sources. However, MOODY'S is not an auditor and cannot in every instance independently verify or validate information received in the rating process or in preparing the Moody's Publications.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability to any person or entity for any indirect, special, consequential, or incidental losses or damages whatsoever arising from or in connection with the information contained herein or the use of or inability to use any such information, even if MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers is advised in advance of the possibility of such losses or damages, including but not limited to: (a) any loss of present or prospective profits or (b) any loss or damage arising where the relevant financial instrument is not the subject of a particular credit rating assigned by MOODY'S.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability for any direct or compensatory losses or damages caused to any person or entity, including but not limited to by any negligence (but excluding fraud, willful misconduct or any other type of liability that, for the avoidance of doubt, by law cannot be excluded) on the part of, or any contingency within or beyond the control of, MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers, arising from or in connection with the information contained herein or the use of or inability to use any such information.

NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY SUCH RATING OR OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.

MIS, a wholly-owned credit rating agency subsidiary of Moody's Corporation ("MCO"), hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by MIS have, prior to assignment of any rating, agreed to pay to MIS for appraisal and rating services rendered by it fees ranging from \$1,500 to approximately \$2,500,000. MCO and MIS also maintain policies and procedures to address the independence of MIS's ratings and rating processes. Information regarding certain affiliations that may exist between directors of MCO and rated entities, and between entities who hold ratings from MIS and have also publicly reported to the SEC an ownership interest in MCO of more than 5%, is posted annually at [www.moody.com](http://www.moody.com) under the heading "Shareholder Relations — Corporate Governance — Director and Shareholder Affiliation Policy."

For Australia only: Any publication into Australia of this document is pursuant to the Australian Financial Services License of MOODY'S affiliate, Moody's Investors Service Pty Limited ABN 61 003 399 657 AFSL 336969 and/or Moody's Analytics Australia Pty Ltd ABN 94 105 136 972 AFSL 383569 (as applicable). This document is intended to be provided only to "wholesale clients" within the meaning of section 761G of the Corporations Act 2001. By continuing to access this document from within Australia, you represent to MOODY'S that you are, or are accessing the document as a representative of, a "wholesale client" and that neither you nor the entity you represent will directly or indirectly disseminate this document or its contents to "retail clients" within the meaning of section 761G of the Corporations Act 2001. MOODY'S credit rating is an opinion as to the creditworthiness of a debt obligation of the issuer, not on the equity securities of the issuer or any form of security that is available to retail clients. It would be dangerous for "retail clients" to make any investment decision based on MOODY'S credit rating. If in doubt you should contact your financial or other professional adviser.