Heavy Models, Light models and Proxy Models

A Working Paper

by The Proxy Model Working Party

24 February 2014

HEAVY MODELS, LIGHT MODELS AND PROXY MODELS

A WORKING PAPER

The Proxy Model Working Party: Christopher Hursey*, Matthew Cocke, Cassandra Hannibal, Parit Jakhria, Iain MacIntyre and Matthew Modisett

ABSTRACT

The use of proxy models within the insurance sector has grown considerably in recent years, particularly in the area of capital management. This growth has been largely driven by the increased demands of a changing regulatory and risk management landscape set against the inability of traditional modelling techniques to keep up.

This paper takes a look at some of the types of proxy model available to practitioners, suggesting a basic framework for "replicating formula" type proxies into which many current proxy models fit. Within this framework, and drawing heavily on recurring themes of complexity, accuracy and, in particular, use of the model, the options available in the design and implementation of a model are discussed as well as the potential impact of the choices made.

Finally, four specific proxy models are discussed in greater detail, two of which are the subject of a case study. This leads to a key result concerning the distinction between risk scenario accuracy and risk distribution accuracy the key driver for risk capital estimation.

CONTACT DETAILS

*Correspondence to: Christopher Hursey, Charasys Limited, Chestnut Lodge, Officers Row, Bramley, Tadley, Hampshire. RG26 5XL. Email: chris.hursey@charasys.co.uk

CONTENTS

1   INTRODUCTION

1.1   Previous Actuarial Research (see Frankland et. al., 2013) suggests that any Actuarial model is necessarily a simplified representation of the real world, and as such there is a huge amount of judgement involved in stripping the "real world" into the relevant components within the constraints of today's computing power and timing requirements.

1.2   The situation has become like the serpent eating its own tail because actuaries must now make models of their own models. Traditional models used for insurer's balance sheet calculations are complex functions of millions of inputs, with perhaps hundreds of stochastic inputs. However, the regulations and management do not stop there. Insurers must derive "Solvency Capital Requirement" which requires them to repeat these calculations under many different scenarios so as to have a high degree of confidence of meeting their "realistic balance sheet" in extreme adverse scenarios. While computational power has increased exponentially over the last several years, it would seem our demand for financial calculations have increased as an exponential of an exponential.

1.3   There are three basic methodologies to meet these demands:

- Vastly enhancing modelling and computing capacity to try and carry out 'stochastic on stochastic' runs.
- Speeding up and optimising models
- Building 'light' or 'proxy' models that that largely replicate the 'heavy' models but run much quicker.

1.4   In this context, a natural separation of modelling has occurred. Some models, termed "heavy models", are developed that best fit reality (within computing constraints) so as to come up with our "realistic" balance sheet based on today's market conditions, but involving lots of different future outcomes. These models cannot be run as often as required so simpler models, termed "proxy models" or "light models" are developed to mimic the heavy models. These light models can then be used to explore more scenarios.

1.5   This paper concentrates on this third approach of 'light' or 'proxy' models. However, as we noted that there were a huge number of choices in moving from the "real world" to a heavy model, there are likewise a large number of choices when moving from a heavy to light model. In fact, it may be better to think of all these possibilities as a spectrum.

1.6   We find that models range from light to heavy with light being the least complex and often the fastest, such as polynomials, and heavy being the most complex and often the slowest, for example a cashflow projection model.

1.7   Where a model lies in the range from light to heavy will often depend on its degree of complexity.

1.8   A proxy model is less accurate than a heavy, detailed model, but is more agile. What are the draw backs in accuracy, and what are the advantages in terms of speed, cost and management information? What types of proxy models are used or considered? How do actuaries present the limitations of each? This paper explores the implications of using such models.

2 BACKGROUND

2.1 A BRIEF HISTORY OF MODELLING TECHNIQUES

2.1.1 In order to provide some context for the discussions that follow we begin by presenting a brief history of modelling methods. This description is by no means intended to be comprehensive and it is also perhaps fair to say that it is more relevant to complex liability models, the setting for much of the work in the area of proxy models.

2.1.2 In the beginning, well perhaps not in the beginning, but within the lifetime of most actuaries, deterministic formulae comprising mainly of commutation functions were in common use. A vector of cash flows and a vector of discount rate were combined via a dot product (known in excel as a sumproduct) to derive a present value. The process was easily generalized to several vectors to include other factors such as lapses.

2.1.3 Influenced by advances in technology, especially the computerised spreadsheet, the use of cashflow models began to gain prominence. The great advantage of cashflow models is that it was easier to model complex systems of cashflows and allow for path dependency. In particular, the previous methodologies did not allow cash flows to vary with discount rates, or for any of the different elements to interact. In particular, interest rates or stock returns in one period might affect lapses or guarantees in a later period, and the notion of path dependence came to be a prominent issue.

2.1.4 Cashflow models allowed evaluation at multiple time-points throughout the projection period, and furthermore allowed evaluation of other statistics of interest, for example, not just of cashflows but also net asset and liability positions. This allowed a wider array of issues to be addressed in the modelling, such as liquidity.

2.1.5 Later still, the regulatory demand for the recognition of options and guarantees coupled with further advances in technology led to the need for, and naturally the development of, stochastic models.

2.1.6 The focus for stochastic models became the evaluation of liabilities and their guarantees at one moment in time, the so-called "Time zero". This can be regarded as a step backward. We regressed from multiple time-point evaluations to a single time-point evaluation of the liabilities, i.e. the deterministic cashflow model providing values at each time-step along a single path was replaced by stochastic cashflow models providing the single stochastic value at time zero derived from the outcome of a large number of paths. Put another way, the capability and ability to perform projections of liability values became more limited because the valuations at a single point in time were so complicated

2.1.7 These models were, and still are, relatively slow. For the more complex insurance liabilities, with-profits in particular, the number of scenarios that can be run in any one valuation exercise remains limited by computational power.

2.2   WHY ARE PROXY MODELS NEEDED?

2.2.1   In summary, influenced by advances in computer technology and modern financial economic theory there has been an overall reduction in the use of analytic functions and greater reliance on cashflow models to provide answers. This development would appear to be due to the relative ease with which a cashflow model can be developed in comparison to an analytic model. The introduction of stochastic modelling has then led to cashflow models being run under many thousands of simulations to provide a single evaluation.

2.2.2   Problems now arise because this evolution was predicated on a requirement for producing only a small number of scenario results. However, over recent years there has been a significant increase in regulatory and risk management demand for information which has in turn led to a large increase in the number of scenario results being requested. At the extreme many thousands of scenario results may be required for a Solvency II internal model.

2.2.3   While computational power has increased dramatically in recent years, the demand for scenario analysis has increased exponentially quicker. Models and infrastructure have been developed through the years to cope with the production of scenario results numbering in the tens, however, scenario results numbering in the hundreds or even thousands are now being demanded. Solvency II in particular has led a number of life and general insurers to develop internal capital models, in which hundreds of thousands of potential scenarios are produced reflecting a range of possible outcomes for economic and insurance risks. Within each of these scenarios, the insurer revalues its balance sheet, and the solvency capital requirement is set so as to ensure solvency in all but a one in two-hundred year event. In other words, the 'tail' of the capital distribution needs to be covered.

2.2.4   While the basic concepts of simulation-driven capital modelling will be familiar to a number of practitioners, the challenge remains as to how to revalue a balance sheet in thousands of different scenarios within a short space of time. The calculation of liabilities itself is a complicated process, and computing capacity is finite. A number of simplifications are needed and the trick is to ensure that the accuracy of the result is not compromised.

2.2.5   As far as cashflow models are concerned, the modelling demand has finally overtaken technological supply. This has led to the introduction of replicating formulae and other proxy models in order to replicate the more complex cashflow models and thus cope with the increased demand.

2.2.6   This is the subject of this paper; the proxy models that are used to bridge the gap between the demand placed on cashflow models and the limited technology providing it.

2.3   WHAT IS A PROXY MODEL?

2.3.1   So what is a proxy model? All models model something; however, it is useful to distinguish between those models which approximate reality and those

which simply approximate a more complex model. The distinction of a proxy model, therefore, is that it models another model.

2.3.2   The primary example of a proxy model arises in capital requirements modelling. A typical company spends considerable effort to run a number of valuations of the company, for discussion we will say 50 valuations under various scenarios. This relatively low number arises because each valuation is quite involved, representing calibration of scenarios across interest rates, equity markets, currencies, lapse assumptions, mortality, and so on. Each of these valuations might be a Monte Carlo valuation involving thousands of simulations, producing a single time-zero value. However, these 50 valuations are not sufficient to deduce a 1-in-200 stress for the company by themselves. The desire is to test many more scenarios, say 10,000 or 100,000. However, the technology does not allow so many different valuations. So a proxy model is developed and employed.

2.3.3   The proxy model is designed to reproduce the 50 valuations, but also provides values for other combinations of the underlying variables. Furthermore, this proxy model can be run quickly. It is often in current practice a polynomial of the underlying variables, although this paper discusses other models. The 10,000 or 100,000 scenarios to calculate the 1-in-200 Value-at-Risk (VaR) stress for the company are run using this proxy model, not whatever process made the original 50 valuations. The 1-in-200 VaR is of particular importance since the Solvency Capital Requirement in Solvency II is based on this.

2.3.4   Notice that the proxy model is calibrated to replicate another, more complicated model. It is not directly a model of reality. In the typical example of a polynomial proxy model, its only claim to be a valuation of the company is that it has the same value for the 50 particular scenarios created by the complex model.

2.3.5   The jargon for the complex model is a "heavy model" or sometimes on valuation will be called a "heavy lift". The proxy model is called a "light model". For example, one says in the above example that 50 heavy lifts were used to calibrate a light model (or proxy model) in order to run 100,000 scenarios.

2.3.6   The above example, while representative of the situation, is not universal. The number of heavy lifts or the number of light model scenarios run may vary. The proxy model may not be a polynomial, rather something else.

2.3.7   For our purposes and the remainder of this paper we define proxy models as those models approximating a more complex model. Going further, however, we can also make the distinction between proxies that attempt to emulate the output of a more complex model and those that attempt to emulate the model itself.

3   DESIGNING AND CHOOSING A MODEL

3.1   TYPES OF MODEL

3.1.1   There are a multitude of choices to be made when building a proxy model. To help us cut through the myriad of options available, we find it useful to think of most proxy models as being a replicating formula consisting of a number of formula elements with each element being allocated a coefficient.

3.1.2   Specifically, suppose we have N risk drivers, $R_1,...,R_N$ which take on the values $r_1(s),...,r_N(s)$ for scenario s and each scenario produces value $y(s)$. We would like to fit this with a proxy function, so we select a number of basis functions of the risk drivers, $X_k(r_1,...,r_N)$ for $k=1,...,K$. The choice of basis function is critical, so let us describe a couple of choices for illustration.

3.1.3   For a polynomial proxy function, the functions X_k are polynomials in the risk drivers $r_1,...,r_N$. This choice is often justified by the Stone-Weierstrass mathematical theorem (Stone, 1948) that if the degree of the polynomials is high enough, then any continuous function can be fitted to an arbitrary degree of accuracy. We challenge this justification of polynomials later (ref. 6.2.4) but offer an alternate justification (ref. 4.2.5 to 4.2.19 & 6.2.64 to 6.2.72) in its place.

3.1.4   For a portfolio replicating model, the functions, X_k, are assumed to be bond-pricing formulas or other security pricing formulas. However, without generalising further, this means that the risk drivers used are restricted to only that subset of drivers $r_1,...,r_N$ that refer to financial markets, such as interest rates, corporates spreads, equity prices, volatilities, currencies, and so forth

3.1.5   It is possible to consider other functions, and a number of alternatives are considered in chapter 6.

3.1.6   No matter which basis functions X_k are selected, the next step is to find the proper combination of these that *best* reproduce the values for each portfolio. In other words, one must solve for $\beta_1,...,\beta_k$ in the following system of equations:

$$\sum_{k=1,...K}\beta_k X_k\big(r_1(s),...,r_N(s)\big)= y(s)$$

For scenarios s=1,…,S where S≥K.

This is often written in matrix form as $\beta X=y$

3.1.7   Where $S>K$ we have an over-determined system for which an exact solution is usually not possible thus requiring a *best* solution be determined (determining the best solution is discussed in section 5.3).

3.1.8   We note that the above system of equations is linear in the $\beta_1,...,\beta_k$ for any choice of the basis functions, $X_1,…,X_K$. A common impression that many

actuaries have is that the linear algebra solves for the proxy model only when the basis functions are polynomials, but as can be seen from the above, the system can be linear for any basis function. By viewing proxy models in this way we are able to identify the fundamental issues and processes common across the different model types and hopefully provide a framework for comparison of different approaches.

3.1.9   We have already commented that when the basis functions, $X_1,\ldots,X_K$, are polynomials, this is the replicating polynomial proxy function which is covered in Section 6.2. When the basis functions are market security functions of market risk drivers then this represents the replicating portfolios proxy model discussed in Section 6.4. The commutation functions methodology (Section 6.5) uses a basis function focused on commutation functions. All three of these fit into the above general description of a proxy model and how it is derived.

3.1.10   The radial basis methodology (described in Section 6.3) is different in that it uses the values of the scenarios themselves as the basis functions and assumes that the proxy function is a function of the "distance" between point considered and the fitting scenarios. The trick with the radial basis methodology is to get the correct weighting between 'local' effects (where the scenarios near the point being considered influence that point's value) and 'global' effects (where the scenarios further from the point considered have an impact).

3.1.11   A methodology of increasing prominence, and therefore worthy of mention, is that of Least Squares Monte-Carlo (LSMC). However, we find that LSMC is not a different type of model at all, often being implemented to calibrate a replicating polynomial. Instead, it is a method characterised by its method of calibration, making use of a large number of scenario results, $y(s)$, the number being large in relation to the number of basis functions, X_k. This methodology is discussed further in section 5.4.

3.1.12   At this point we should emphasise that not all proxy models would neatly fit into the linear solving-for-weights paradigm describe above, as there can be infinitely many ways of simplifying a heavy model. However, we believe that this forms a valuable framework for comparison, and that many prevalent proxy models in the industry do indeed fall within this framework. Thinking in terms of this paradigm will certainly help the reader in the sections that follow where we discuss, in a generalised sense, some fundamental choices being made in choosing a model.

3.2   USE OF THE MODEL

3.2.1   The first choice we consider here is the uses to which the model will be put.

3.2.2   Initially, any heavy model (designed to emulate reality as best as it can) will have been developed for a specific purpose or to carry out a specific role. However, the model will often be subject to subsequent developments in order to fulfil additional roles or for other purposes. Whilst this is usually possible

for the primary heavy models, it may not be so straightforward for some proxy models, as many fundamental design choices are made early in the process.

3.2.3 By their nature, a proxy model is a less complex version of some other model and this loss in complexity is usually accompanied by a loss in the ability to reproduce some of the behaviours of that which it is trying to emulate. However, rather than a smooth loss in ability across the whole model, a proxy model will often sacrifice some behaviours altogether in favour of other behaviours.

3.2.4 In other words, in order for a proxy model to do one thing (almost as well as the original) it must lose the ability to do something else. This will often be required so as to maintain the proxy model's value in terms of speed and accuracy.

3.2.5 It is therefore important that the intended use of a proxy model is considered before choosing, designing and then building the model since subsequent attempts to adapt it may not be possible or may be at the expense of the original purpose. Furthermore, the model may be used for purposes for which it is simply not suitable.

3.2.6 It is in this context that we give particular consideration to the use of proxy models in capital measurement and management. Proxy models are used in capital management to provide a proxy full distribution from which appropriate percentile results can be drawn (such as the ubiquitous 99.5th percentile or 1-in-200 as used by both the UK Individual Capital Assessment and the European Solvency II capital regimes).

3.2.7 Despite a primary interest in the capital distribution, it is often the individual scenario results that draw most attention. This is due in part to the fact that a comparison between primary and proxy scenario results is often the only way of assessing accuracy of the proxy. However, there is also the temptation to use the multitude of scenario results for more detailed capital analysis and management.

3.2.8 It is here that care must be taken as some proxy models may be ill-suited to this use, being very inaccurate at the individual scenario level. However, an important result is that a model need not be accurate at the scenario level for it to provide an accurate description of the capital distribution and likewise an accurate assessment of required capital. This result is discussed in greater detail in section 4.2 when considering the goodness of fit of proxy models and again in section 6.2 when a case study involving replicating polynomials is performed.

3.3   COMPLEXITY VERSUS ACCURACY

3.3.1 Having briefly considered complexity in the context of use, we now turn our attention to the choice of complexity versus accuracy. If we consider complexity and its association with accuracy, we can increase the complexity of a proxy model by either increasing the complexity of those formula elements, $X_1,\ldots,X_K$, introduced in section 3.1 or by increasing the number of

elements, *K*. And since we normally associate increasing complexity with increasing accuracy, more sophisticated formula elements or an increased number of formula elements will often be associated with greater accuracy. However, this will often lead to slower runtime hence the expected relationship between greater accuracy and slower runtime.

3.3.2 Generally, as each formula element, X_k, becomes more sophisticated and able to capture more complex behaviour, the number of elements required for a given level of accuracy should fall. The important implication of this is that the number of calibration points should likewise fall.

3.3.4 We therefore often have a trade-off between complexity of formula elements and ease of calibration, i.e. for a given level of accuracy, as the complexity of the formula elements reduce, the number of required elements increase and the model becomes more difficult to calibrate. This is an important consideration when choosing a model and shows how the choice between accuracy and complexity will not only influence results but the implementation process as well.

3.3.5 We discuss the impact of this choice in more detail in section 4 where we describe criteria for developing and implementing a proxy model.

3.4   INTUITION

3.4.1 The final choice we consider here, and perhaps a less obvious one, is how intuitive we want the model to be. This may be manifested in terms of formula structure (which is discussed in section 5.2).

3.4.2 At the one extreme, one could be agnostic to intuition and prioritise the descriptive power of the model. This may lead to the use of polynomial fitting, which provides tight bounds within a given range, but where the coefficients (and in particular changes from one year to next) are not intuitive. Thus, for example if the coefficient of x^2y changed from -652 to 1,456 from one year to next, this information in itself may not necessarily prove insightful, even though the overall polynomial fits can be shown to fit well over a given range.

3.4.3 On the other hand, one may try and use a more intuitive formula structure such as a portfolio of financial instruments / options or commutation functions, each component of which may have an intuitive meaning. For example, a set of with profit liabilities may be represented by a series of portfolio put options, with each option roughly corresponding to a block of business sharing similar characteristics and maturing in the same year. In this case, the movement of the coefficients of the respective options from year to year do provide valuable additional insight.

3.4.5 The two cases above are perhaps extreme examples, but they do serve to illustrate one of the key choices available when designing the model. In an era where a wider range of business functions are expected to use and understand capital calculations, intuitive methods can ease embedding of capital metrics into business as usual processes, as the components have meaning and the coefficients provide insight. Intuitive methods are perhaps more powerful

when there is sufficient knowledge of the problem at hand (i.e. knowledge of the products and liabilities being analysed), as well as resources to design a neat (but potentially more complex) formula structure. On the other hand, when approaching a problem with little knowledge about the products / liabilities, a polynomial method can be used as a general (albeit brute force) method, that would approximately work in a large number of situations.

3.4.6   Perhaps another advantage of intuitive methods is their behaviour outside of their fitting points. A replicating portfolio can be expected to behave broadly sensibly outside of its 'reliable range', whereas a classic criticism of pure descriptive methods such as polynomial fitting is that it may run through all the known points but vary widely between these points (interpolation), or diverge from expectations outside the fitted range of these points (extrapolation).

4   METHODS FOR EVALUATING A PROXY MODEL

4.1   INTRODUCTION

4.1.1   In order to make an informed choice of model we need means of evaluating the various models; some objective, some not so.

4.1.2   The choice of model should be influenced primarily by the uses to which the model will be put and should therefore be evaluated in this context, i.e. how well does the model achieve what it was initially designed to do? In the previous section we discussed in general terms some issues to consider in choosing a model. It is useful to expand on that discussion and address aspects of accuracy and complexity in more detail in order to develop criteria for assessing a model.

4.1.3   We start by discussing accuracy.

4.2   QUALITY OF FIT

4.2.1   A variety of statistical methods can be used to assess the quality of fit of the proxy model. Many of these are well known and often used for comparing the results from a known quantity, such as heavy model results on out-of-sample test points, to an estimated quantity, such as proxy model results. These include, but are not limited to:

- Anderson-Darling
- Kolmogorov-Smirnov
- Cramer Von Mises
- Shapiro Wilks
- Chi squared
- Akaike Information Criterion
- Bayes Information Criterion
- QQ plots
- PP plots
- R squared (regression)

4.2.2   The first five in the list are statistical tests used to assess whether a given distribution is an appropriate representation of the observed data. After applying this test there may be more than one possible 'answer'. Information criterion such as Akaike and Bayes are used to assess the trade-off between complexity of the formula structure and the goodness of fit so can help to shorten a short list. QQ and PP plots are visual aids used to assess the goodness of fits across quantiles or percentiles and are particularly useful when looking at specific regions of the distribution. This is different from the R squared technique which considers the entire distribution.

4.2.3   In the context of proxy models there are different points to consider when applying these tests as, in general, attaining a good quality of fit across all aspects of the calculation is unlikely without compromising the speed and convenience of the proxy model. Points to consider include:

- Distribution accuracy – is accuracy required across the entire distribution of each component, or only in particular regions?
- Scenario accuracy – is accuracy required for some subcomponent of the calculation or for the entire calculation? For example in the capital requirement calculations accuracy particular focus is placed on specific scenarios (e.g. 'the' 1-in-200 scenario) but in other cases the distribution of the capital requirement may be the focus.
- Component accuracy – is accuracy required for particular factors or for the combined result?

4.2.4   Answering these questions is required to assess the quality of fit of the model, but in order to answer these questions the uses of the model must be considered. Before considering some examples, we return to the subject of distribution accuracy versus scenario accuracy introduced in section 3.2.

Distribution Accuracy versus Scenario Accuracy

4.2.5   As already discussed, the growing interest in proxy models has largely been driven by the need to run many thousands of scenarios from which the capital distribution can be estimated and quantile results drawn. An issue arises in that an assessment of distribution accuracy will necessarily be based on scenario accuracy, itself being based on a limited number of out-of-sample test scenarios. An important result, however, is that distribution accuracy, and therefore required capital assessment, is not necessarily dependant on scenario accuracy.

4.2.6   This is best illustrated with an example drawn from our replicating polynomial case study (ref. 6.2). Consider the chart in Figure 4.2.1 showing error percentages from 20,000 out-of-sample tests of a proxy cost-of-guarantee model.

*Fig. 4.2.1 – Proxy 'Cost-of-guarantee' error percentages*
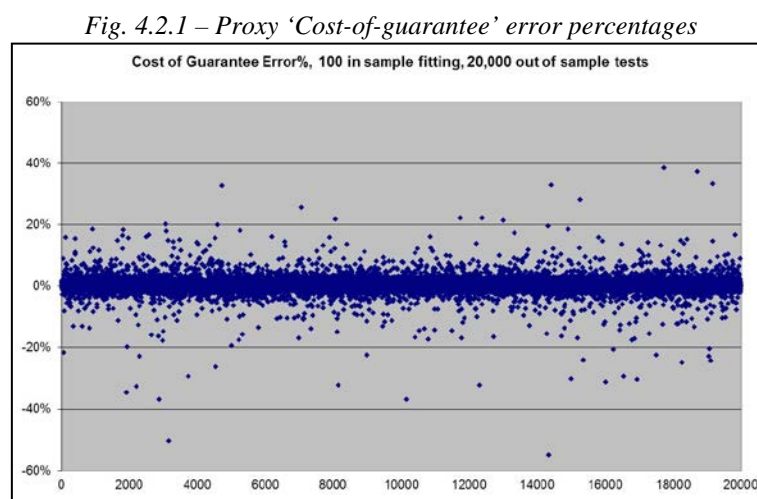
4.2.7   By most conventional measures, in which actual results are compared to proxy results for individual scenarios, the results demonstrate a poor quality of fit,

ranging from -55% to +38%. (This fit could have been improved using a variety of methods but this is not the purpose of the exercise here.)

4.2.8   Surprisingly though, we found the quantile results to be very accurate across the whole distribution. In particular the error at the 99.5th percentile was less than 0.2%. This is demonstrated in figure 4.2.2 in which the ranked errors from the chart in figure 4.2.1 (distribution of scenario errors) are compared to the error in ranked results (distribution error given by the difference between ranked proxy results and ranked actual).

Fig. 4.2.2 – Ranked errors vs error in ranked results
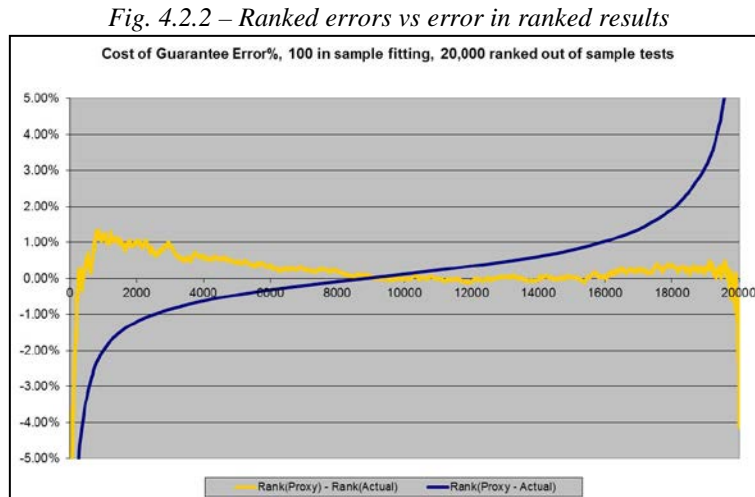
4.2.9   So how can the scenario results be so inaccurate and yet the quantile results be so accurate?   Based on our experience of approximations in one risk dimension, it is tempting to assume that proxy model errors increase as scenarios become more extreme. Figure 4.2.3 shows the error curve resulting from just such an approximation; in this example an optimal quadratic polynomial approximation function.

*Fig. 4.2.3 Error curve in one risk dimension*
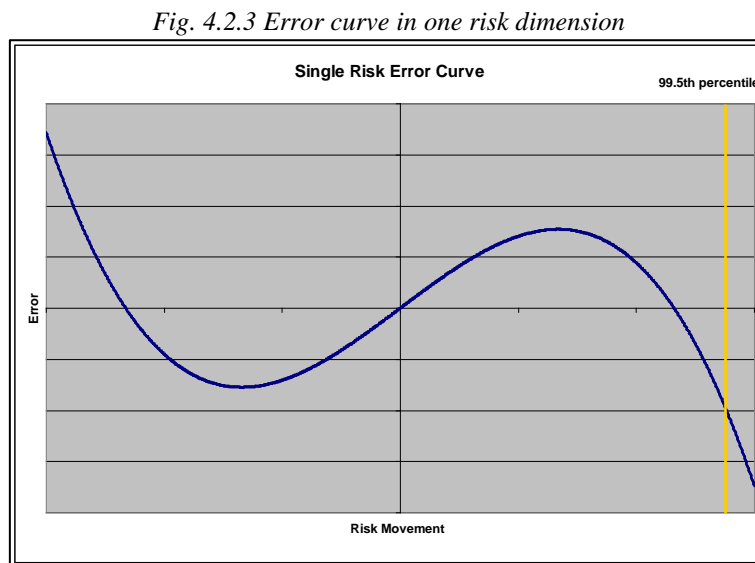
4.2.10 Beyond a certain point there are no further turning points, and errors continue to increase in magnitude. Drawing quantile results in the tails leads to

increasingly large errors as the scenario being considered becomes more extreme.

4.2.11   In fact, this occurs not due to the extremity of the event but due to error bias in the region of interest. All errors are of one sign and for any particular percentile there is only one point to choose. Therefore, the error in the quantile result is dictated by the error in that scenario result.

4.2.12   However, in multiple risk dimensions the single point is replaced by the curve of constant loss. This is a curve representing the combinations of different risk variable values that all give the same result. For example, if equity values drop by 10% and lapse rates increase by 5%, we might get the same answer as if equity values rise by 5% and lapse rates fall by 10%, and in this case these two scenarios would be on the same curve. If we could plot the errors along this path we would not expect the errors to be of one sign, the proxy being greater or less than actual at different points along the path. Figure 4.2.4 illustrates.

Fig. 4.2.4 – Curve of constant loss
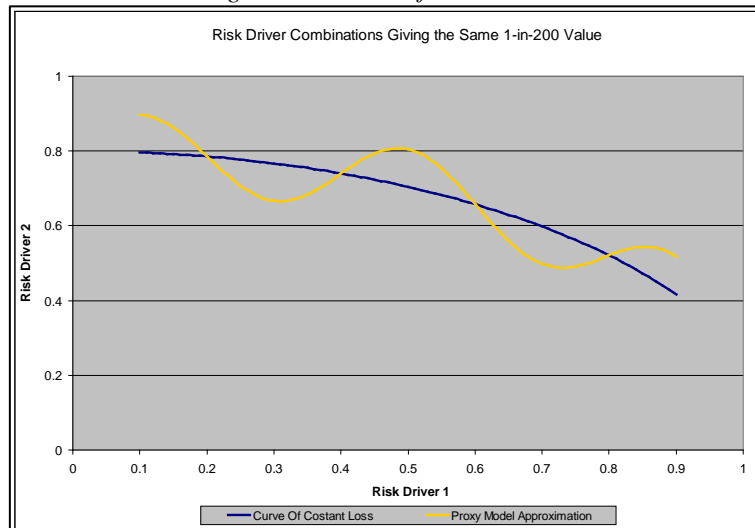
4.2.13   At specific points the errors can be large, up to 60% in our example, but the nature of least squares, which does a good job of minimising average error, removes some of the error bias along the path of constant loss. However, this is by no means guaranteed, even with a least-squares fit.

4.2.14   In order for the result to apply we wish to minimise error bias along the curve of constant loss. More formally, we wish to have an expectation of error equal to zero.

4.2.15   So what does this mean for us in the real world? Returning to our example, comparing ranked results of actual versus proxy we find that the scenario numbers in each of the two lists of results do not match. In particular, the biting scenario in the proxy model, that scenario providing the 99.5[th] percentile result, is different from the biting scenario in the heavy model that provides the actual result. However, despite the biting scenarios being different, the results of the two scenarios are very similar. This is illustrated in

table 4.1 which compares the biting scenario results from each of the proxy and heavy models.

Table 4.2.1

Proxy (£m)	Actual (£m)	Error (£m)	Error (%)
441.5	442.2	(1.1)	(0.17%)

4.2.17 In reality we are unlikely to have the 'actual' quantile result so may be inclined to check our proxy result by running the implied biting scenario through the heavy model. Table 4.2 shows the results when the biting scenario given by the proxy model is run through the heavy model.

Table 4.2.2

Proxy (£m)	Actual (£m)	Error (£m)	Error (%)
441.5	448.3	6.8	1.52%

4.2.18 From here on, the inaccuracy at the biting scenario can dominate proceedings if we are not careful. In particular, it may be concluded that the correct result is £448m, overstating the true result of £442m. Worse still, it may be decided to change the calibration of the proxy model to provide a better fit at the proxy model derived biting scenario.

4.2.19 We conclude, therefore, that using an inaccurate model to determine a 'biting scenario' which is then subjected to more detailed analysis may not be appropriate. The above results show that the capital measured by a proxy may be relatively accurate even if the scenario producing it is not. It also raises the issue of just how much value there is in analysing a single biting scenario. If the proxy model has already provided the correct capital result then from a risk management perspective it may be more appropriate to test a range of biting scenarios along the curve of constant loss in our primary models.

4.2.20 It is also instructive to consider the problem of estimating VaR from another viewpoint. In determining VaR through a proxy model, there are two possible sources of error:

- **Proxy model error**, where the value in a particular scenario differs to the value from the heavy model in that scenario,
- **Stochastic error**, which reflects limitations in Monte Carlo simulation. The estimated VaR through Monte Carlo simulation will not equal the true VaR. (There is discussion of simulation error in Frankland et. al., 2013.)

4.2.21 It is interesting to consider research which investigates the relative size of the two error sources. This is considered in a slightly different context by Stentoft (2004), looking at pricing American options through a regression that uses a Least Squares Monte Carlo approach. Under this approach, the option value before maturity, assuming it is not exercised early, is estimated through a value determined by polynomial regression. This is compared to the value under early exercise to determine the option value in the scenario considered. Under this approach, the two sources of error are as in 4.2.20 – proxy model

error is error caused by the polynomial approximation; and stochastic error is caused by an insufficient number of simulations in the Monte Carlo projection.

4.2.22 Stentoft found that the number of simulations in this setting is a more important driver to accuracy than the number of terms in the polynomial, proving that the estimated option price converges to the true option price under certain conditions. In particular, convergence is guaranteed under various assumptions provided that:

$$\frac{M^3}{N} \to 0 \; as \; N \to \infty \; and \; M \to \infty$$

Where M is the number of terms taken in the Legendre polynomial and N is the number of simulations. Hence the number of terms in the polynomial can increase at a materially slower rate than the number of simulations.

4.2.23 We believe a similar result applies in the setting of this paper – using estimated function values in a Monte Carlo simulation to determine the VaR of the true function under some statistical distribution. [The details will appear in Appendix 1.]

4.2.24 Given this discussion around the distinction between scenario accuracy and distribution accuracy, we now provide some example uses of the model and their respective accuracy considerations.

Example Uses of the Model

Daily reporting
4.2.25 A proxy model used for frequent reporting such as daily capital calculations would most normally encounter small movements; although in the event of a sudden large movement should also provide reasonable results. If the purpose of the daily reporting is to provide fast and accurate results in the event of a shock then when assessing quality of fit, we may focus on the areas around the current scenario. However, since the model may be used to assess capital requirements, the tails of the distribution are potentially also important.

4.2.26 This suggests that calibrating across the entire risk distribution is important; near the base case is required to assess the impact of small market movements and the tails are required to estimate capital requirements. Also as individual scenarios (univariate stresses) are unlikely to be assessed, more accuracy is required on the combined result.

Stress testing
4.2.27 Stress testing involves changing one single component and assessing the impact on the combined result. So when assessing the quality of fit we may focus on the goodness of fit of particular components. As stress tests by their very nature consider extreme results the fit of the whole distribution should be assessed.

Scenario testing

4.2.28 Scenario testing involves changes combinations of components and assessing the impact on the combined result. In this case when assessing the quality of fit the focus is on combined results rather than particular factors and again as extreme events are investigated, across the distribution.

Setting limits or appetites

4.2.29 Appetites, although driven by the combined result, are commonly set at the component level rather than the combined result level. This is largely because prescribed actions used to manage against those limits are more easily prescribed at the component level. This implies the focus of any testing is on individual factors. However, the findings discussed above under distribution accuracy versus scenario accuracy (ref 4.2.19) suggest a range of biting scenarios should be considered when deriving appetites or limits.

Strategic analysis

4.2.30 This "what if" analysis is not prescribed and can involve any number of changes including adding or removing entire components. Changes like these will often change interactions between components and are notoriously difficult for proxy models. Initially the focus is usually the combined result but it is likely that components will also be analysed.

4.2.31 In summary, the regulatory and risk management environment encourage businesses to be able to carry out calculations for different purposes with increasing frequency, speed and accuracy. The five examples given are just a small sample of the uses of proxy models. Consequently the actual use for a single proxy model usually includes all of the above examples and more, and so the relative importance of the goodness of fit tests requires significant expert judgement.

4.3   EASE OF IMPLEMENTATION & COST

4.3.1 There are then the related issues of ease of implementation, cost and speed of implementation. Solvency 2 requirements provide a defined reporting timeline to aim at but there is also increasing demand for faster model estimates for firms looking to embed economic capital metrics as part of an Enterprise Risk Management framework. Set against the ever increasing demand for faster model estimates is the need to consider the quality of fit across a range of metrics supporting regulatory reporting, internal risk appetite and capital target reporting and stress and scenario testing.

4.3.2 The competing challenges of speed and quality will drive choices on implementation alongside the cost of initial implementation and ongoing maintenance.

Process Design

4.3.3 Frequency of the reporting cycle and the length of the reporting window will influence choices around the design of the reporting process and timing of calibration activity. For example, for a company with a simple balance sheet and efficient heavy models it may be feasible to carry out calibration as part of a single

reporting process. For more complex balance sheets, it will be necessary to calibrate the proxy model in advance of the reporting period and use roll forward techniques along with a reduced volume of modelling required to "validate and true up" as part of the reporting process. Where there are multiple uses for the proxy model different levels of validation may evolve. For example:

- Daily or monthly solvency monitor estimates driven by proxy model with a trigger framework defining when adjustment would be required.
- Quarterly regulatory reporting with re-calibration where practical and a defined validation and adjustment process.

Calibration

4.3.4   Consideration should be given to the following factors when choosing the calibration approach:

- Use(s) of the model
- Quality of fit and error tolerances
- Model stability

Calibration Method
4.3.5   The choice of calibration technique will be influenced by the quality of fit required and the uses to which the model is to be put. Where there are multiple uses demanding a good fit across a range of metrics efficient calibration techniques allowing the calibration of more complex proxy functions are likely to be required.

Calibration Frequency
4.3.6   The frequency of calibration activity will be driven by the stability of the model (discussed below) and the extent to which any management actions and changes in risk profile over time can be adequately captured without the need for re-calibration.

4.3.7   Where management actions or evolution of the risk profile are significant factors this is likely to point to more frequent re-calibration in turn influencing the spend on initial implementation.

Initial Implementation

4.3.8   Typically the calibration phase might demand a high volume of fitting points to be generated using the heavy models. Efficient calibration techniques such as Least Squares Monte Carlo may require fundamental changes to the way heavy models are set up and run to support a large number of fitting points relative to more traditional curve fitting approaches. In addition to model changes, these techniques may also change the way in which heavy model output is validated with the need to place more focus on automation and validation of model set up.[1]

---

[1] Automation is arguably important for traditional curve fitting as well. If say, 200 fitting points and 100 out-of-sample test points are required, with 5000 simulations for each point, then automation may be critical to run the process in a reasonable time-frame.

4.3.9   Alongside costs to implement the proxy model tools themselves this means that the cost of initial implementation can be significant. Higher spend at outset needs to be set against reduced ongoing costs obtained by the increased speed of the calibration and proxy modelling processes. However, the marginal benefits of additional model spend will start to reduce as process hotspots and validation activity start to impact the critical path.

4.4   MODEL STABILITY AND FACTORS THAT INFLUENCE IT

4.4.1   For any model, stability is an important issue. This generally means that the output of the model does not change significantly for small changes in the inputs.

4.4.2   An example for readers familiar with Monte Carlo simulation would be the requirement that insignificant changes in the random number generation should not impact the final result. This requirement is normally met by performing more simulations so that the significance of any random number is minimized. For proxy models that are used for Monte Carlo simulations, this would still be a requirement, though there are additional requirements related to a proxy model.

4.4.3   Stability requirements specific to a proxy model are:

- Small changes in heavy run results (inputs to the proxy model calibration) should not cause large changes in the proxy model itself. Do small changes in heavy run results create large differences in light run results?
- The proxy model should be stable over time. For example, if a proxy model is to be recalibrated annually but used quarterly, then the methodology employed should be one in which quarterly re-calibrations would have been stable if they had been employed.

4.4.4   In general, we test stability by perturbing the inputs and confirming that the change in the outputs is not significant.

4.5   COMPLEXITY – MANAGEMENT ACCEPTANCE

4.5.1   Complexity is a common theme across many of the factors that influence model choice and design, not least in terms of accuracy and cost. However, the ability to understand, interpret and then communicate the results produced by a proxy model should not be underestimated as a standalone issue to consider.

4.5.2   The risks of putting complete faith in a model without understanding its weaknesses and limitations, the so-called 'black box', are well documented. At the other end of the scale, however, there are the risks associated with a lack of faith in, and management acceptance of, a model.

4.5.3   In particular, management acceptance of a proxy model will impact the degree to which the information it provides will be trusted and therefore heeded. Modern risk management demands much more of models than simple measurement and data production. The data produced by a model needs to be

converted into usable information before management will be able to use it to inform and guide decision making. The ease with which this conversion can take place will be influenced by both the complexity of the model and the level of intuitive understanding it provides.

4.5.4 Often, but not always, the complexity of a proxy model will be closely associated with the level of intuitive understanding it provides; the simpler a model becomes, and thus the easier it is to understand, the less relation it might bear to reality, perhaps making it more difficult to have faith in the output. This is made worse if it proves difficult to interpret and explain unexpected results due to the lack of real-world meaning associated with some or all of the model's components. This is a criticism that some may level at replicating polynomials.

4.5.5 Conversely, the greater intuition provided by a model that is more reflective of reality may come at greater cost due to increased complexity. The output may be easier to interpret but then the model is more difficult to understand. However, the very reason for needing a proxy model in the first place is to reproduce the output of a more complex model. Therefore, as the proxy becomes more complex, it moves closer to that which it is meant to approximate and becomes of less value. A balance needs to be struck between a model that is a sufficient simplification of a more complex model so as to have value whilst remaining sufficiently complex to be able to provide meaningful and accurate management information.

4.5.6 The heavy-lift cashflow models are an example of very complex models attempting to match reality as closely as possible and in which there is a high level of trust in the output, even by those that do not know the inner workings of the model. At the other end of the spectrum, there remains a degree of uncertainty as to the level of reliance that can be placed on proxy models. Admittedly this contrast arises not only from the different level of intuitiveness of the models but also from the different lengths of time over which the models have had to become embedded in operations.

4.5.7 Ultimately, in order for senior management to trust the model, they either need to trust the information it provides, understand the model themselves, or a combination of both. Thus in the context of gaining management acceptance, one could view the choice as being between a model that is easy to understand or results that are easy to understand (interpret). Obviously this is a generalisation and arguably a little too simplistic since the reality will be a balance between the two. But, it does serve to illustrate more succinctly the choice that needs to be made.

5   CALIBRATION

5.1   INTRODUCTION

5.1.1   Once a model basis has been chosen, it will need to be calibrated. Using our replicating formula paradigm introduced in section 3.1 we recognise two separate stages in the design and calibration of the formula common across most model types.

5.1.2   The first stage is determination of the formula structure, deciding the elements, X_k, to be included in the formula. For example, in respect of a replicating polynomial it is deciding whether to include in the formula an x^2, an xy^3 or an xyz etc. In respect of a replicating portfolio it is deciding which assets are included in the final replicating portfolio.

5.1.3   Once the formula structure and the included elements have been determined, the second stage of calibration is determining the coefficients, β_k, of each element.

5.1.4   Both these stages may occur at the same time, more commonly where an automated computer algorithm is being utilised, or the two stages may remain distinct processes.

5.1.5   Whether the two stages are run as separate processes or not, their separate identification is useful so as to recognise the differing objectives fulfilled by each stage. When determining formula structure the objective is to build a model that can reproduce the behaviour of a more complex model whilst the objective when determining the coefficients is to reproduce the results of a specific dataset.

5.1.6   Before considering those objectives further, and in the context of design and calibration, it is also worth considering the two different environments in which the chosen model will be operated. The development environment, in which the model will be designed, built and tested, invariably involves both stages of the calibration process, often in an iterative refinement process. Once in a production environment, however, the model will often be required to produce results to restricted time-scales. As such, once a model is in a production environment, calibration may be limited to the second stage only, determining formula coefficients.

5.1.7   Within both stages there are various choices over methodology to consider. These and other issues are considered in further detail in the following sections.

5.2   DETERMINING FORMULA STRUCTURE

5.2.1   The key objective when determining formula structure is to construct a model that can adequately reproduce the behaviour of a more complex model when subjected to variation in a number of different risk parameters.

5.2.2   Care must be taken here to distinguish between reproducing the behaviour of the underlying model and the behaviour of the results it produces. If relying on data to represent the behaviour of the model then any behaviours present in the model but not represented within the data will likely not be captured by the proxy model. The proxy may fail to be predictive.

5.2.3   Alternatively, relying solely on an understanding of the underlying model may also lead to an inadequate proxy due to limitations in knowledge of the model. A model will invariably have many variable risk parameters and the interactions and interdependencies between those parameters are many and complex. Even where interactions are recognised it may be difficult to predict the nature of those interactions without reference to data.

5.2.4   In reality it is unlikely that reliance would be placed solely on an understanding of the model when trying to construct a proxy and the point is made here merely to illustrate the extreme alternative to relying solely on data. And yet, it is interesting to note that reliance solely on data for the construction and calibration of a proxy model is not unusual.

5.2.5   In order to provide a more intuitive model (even for polynomials) some knowledge of that which we are trying to approximate can be employed alongside data analysis to provide insights into the complex interactions between the risk parameters. Knowledge and data can be used to identify the various risk interactions whilst data is then used to determine the nature of those interactions.

5.2.6   One such approach to building a proxy model may begin with the construction of the formula or components of the formula based on knowledge and expectations for the interactions between risk parameters. Following calibration and testing, the model can be refined based on observed results, i.e. goodness of fit. A series of refine and retest processes may be required before an adequate formula structure is derived.

5.2.7   An alternative is the use of an automated algorithm which will seek to derive the formula structure by testing many different structures against a given dataset, selecting the structure which, following calibration, provides the best overall fit to the data.

5.2.8   However, even where an automated algorithm is used, closer examination reveals that the method is often just emulating a manual process, replacing subjective decision making with objective decision making techniques. The obvious advantage of objective methods, however, is that they can be codified and carried out by a computer. This makes it possible to test a far greater number of formula structures.

5.2.9   Whilst this approach may provide a better fit there is the risk that that fit has been achieved through the introduction of non-existent or inappropriate risk parameter relationships. This can lead to unexpected results when the proxy is used to extrapolate results from the calibration data. That being said, the

failure in this example is not in the use of automation but would be due to a limitation in the decision making algorithm.

5.2.10 Ultimately, therefore, the choice of process for determining formula structure will fundamentally involve a choice between the level of subjective 'expert judgement' and the level of objective 'automation'. Here, it should be made clear that 'automation' can include processes carried out manually. We are referring more to the decision making process rather than the means by which decisions are implemented.

5.2.11 This choice, and the balance struck between the two, will influence to what degree, if any, a model will be predictive as well as impacting on the stability and intuitiveness of results.

5.2.12 If we recall from section 3.4 when discussing intuition and predictive or descriptive models, a predictive model arguably demands a more subjective approach where the model is constructed from pre-conceived ideas about how the various parameters inter-react before being tested and refined. This approach should hopefully provide a more intuitive and predictive model although the degree of truth in this will be influenced by the basis function used.

5.2.13 For example, a carefully constructed polynomial may offer some intuitive understanding to those very familiar with the model but trying to explain why the coefficient of equity squared multiplied by lapses has changed from positive to negative exposes limitations to this intuitive understanding, even for those closest to the model's construction. On the positive side, however, building the proxy model subjectively should at least explain why this formula element is required in the first place, even if the meaning of its coefficient is unclear.

5.2.14 That said, there are significant advantages to using more objective or automated methods, not least in speed and ease of replication, which will themselves have lower cost implications. Without getting drawn too much into a discussion around Actuarial philosophy, it may also be tempting to use a method which utilises a minimum of expert judgement. However, whilst responsibility for design is taken away from those running the proxy model, the same cannot be said of the blame should the model fail. It is vital that the limitations of the model and the point at which it fails are understood and to this end some understanding of the formula structure, however it is derived, is crucial.

5.3   DETERMINING FORMULA COEFFICIENTS

5.3.1 The second stage of calibration is the determination of the coefficients, or weights, β_k, to be assigned to each formula element.

5.3.2 Whilst the objective of the first stage is to build a model capable of emulating the behaviour of a more complex model, the objective of the second stage is to reproduce a specific dataset, or set of outputs, of that model as closely as

possible. This dataset usually consists of a set of 'in-sample' scenario results[2] to which the proxy will be calibrated. A further dataset consisting of 'out-of-sample' scenario results will then be used for testing.

5.3.3 In many ways the second stage will be more straightforward than the first, eschewing subjectivity in favour of objective or mechanical processes and calculations. Even so, there remain a number of choices to be made in respect of the type of calculation process to be employed. These choices will impact the nature and quality of fit of the model so need to be considered alongside the use to which the model is being put.

<u>Target Calibration</u>

5.3.4 The first choice we consider is that of target calibration. By target calibration we mean that metric which we wish to optimise. The most prevalent method (often contracted and referred to as "Least Squares") seeks optimal calibration of the proxy by minimising a single metric, the sum of squared errors. There are alternative targets, in particular the "minimax" problem which seeks to minimise the maximum error, but we limit our attention in this paper to the method of least squares.

5.3.5 Recall the system of equations from 3.1:

$$\sum_{k=1,\ldots K}\beta_k X_k\big(r_1(s),\ldots,r_N(s)\big)= y(s)$$

For scenarios s=1,…,S where S$\geq$K.

This is often written in matrix form as $\beta X = y$.

For S>K, the problem is one of regression for which an exact solution may not be possible. The least squares solution is then found by minimising the function S given by:

$$S(\beta) = \|y - X\beta\|^2$$

5.3.6 Provided that the *K* columns of the matrix *X* are linearly independent, this minimization problem has a unique solution, the formula coefficients being given by the vector β:

$$\beta = (X^T X)^{-1} X^T y$$

5.3.7 Thus the formula is complete, having coefficients for each formula term. Subject to the results of testing on an out-of-sample dataset, the formula can then be used in production on alternative datasets. Before being used in a production environment, however, and as already noted, this part of the

---

[2] Different practitioners will have given the in-sample scenarios various names; calibration scenarios, calibration nodes, fitting points. However, they all refer to the same thing.

calibration process may be repeated many times when determining formula structure.

5.3.8 As noted in section 5.2, subjective or objective decision will factor in the determination of formula structure. However, the decision will often correspond directly to the method of determining formula coefficients, e.g. for a given formula structure the aim is to minimise the sum of square errors and the chosen formula structure will be that which provides the lowest sum squared errors (often implemented as the square root of sum squared errors). Other considerations may factor and criteria set accordingly.

5.3.9 In 5.2.8 we introduced the idea that automation is often just the codification of a manual process, replacing subjective decisions with objective decisions. Taking this a step further we realise that many subjective decisions can be analysed and broken down into objective components and that many decisions, which would be considered subjective, are simply the unconscious application of weights to different outcomes. In this context, we give more importance to some areas of the results distribution and disregard others. This can sometimes lead to the least-squares metric being over-ruled if, for example, a part of the result distribution has very large error despite having lowest sum squared error.

5.3.10 Realising this, weighted least squares can provide a means by which subjective decisions can be potentially codified and automated. In fact, the least squares problem can be generalised further to that of weighted least squares, whereby a weight function is applied to the squared errors.

$$w(s)\left(y(s) - \sum_{k=1,\dots K}\beta_k X_k\big(r_1(s),\dots,r_N(s)\big)\right)^2 = 0, \quad s = 1,\dots,S, \quad S \geq K$$

This can be written in matrix form as $W^{1/2}(y\text{-}X\beta)$, where W is the diagonal matrix of weights $w(s)$ for each scenario $s=1,\dots,S$.

The least squares solution is then found by minimising the function S given by:

$$S(\beta) = \left\| W^{\frac{1}{2}}(y - X\beta)\right\|^2$$

The formula coefficients thus being given by:

$$\beta = (X^T W X)^{-1} X^T W y$$

5.3.11 When the weight function, $w(s)$, is a constant, W is the identity matrix and this simplifies to the traditional unweighted least squares problem. However, the result will only be an unweighted least squares fit provided the calibration scenarios have been drawn from a uniform distribution.

5.3.12 If say, an unweighted least squares calibration is performed on scenarios drawn from a Normal distribution then this is equivalent to performing a weighted least squares calibration on uniformly distributed scenarios, where the weight function is the Normal probability distribution function. An example of this is demonstrated in the Appendix.

5.3.13 Care must be taken, therefore, that if an unweighted least squares fit is required, the calibration scenarios must be drawn from the uniform distribution since drawing the scenarios from anything other than a uniform distribution will result in a sub-optimal fit in the traditional (unweighted) sense.

Regression fitting vs precise interpolation

5.3.14 We have so far considered only regression fitting. Here we have more scenario results, S, than formula terms, K. However, the quality of fit may be poor if S is not sufficiently large in comparison to K. This may lead to a large number of scenario results being required. As such, it may become necessary to consider techniques for improving the efficiency of the process. One way to improve efficiency is to use many inaccurate scenario results as employed by Least Squares Monte Carlo (we discuss this method in section. 5.4). An alternative is to use the minimum number of scenario results possible, S=K, and solve by precise interpolation.

5.3.15 The problem with precise interpolation is that the resulting fit is entirely dependent on, and very sensitive to, the selected calibration scenarios. However, the regression fit achievable through using a large number of scenarios can be emulated using precise interpolation if the interpolation scenarios are picked in a certain way. In the case of polynomial basis functions, Hursey & Scott (2012) showed that, for a given formula structure, by selecting scenarios derived from the roots of Legendre polynomials, a best estimate of the best possible fit can be achieved through precise interpolation.

5.3.16 More precisely, interpolation using Legendre derived scenarios emulates an *unweighted* least square regression fit. This relates to the fact that the Legendre polynomials are orthogonal with respect to a uniform distribution or a weight function, $w = c$, where c is a constant. In fact, using scenarios derived from the roots of polynomials that are orthogonal to other probability distributions, leads to a precise interpolation fit that emulates the weighted least squares regression fit, where the weight function is the relevant probability distribution function. For example, using interpolation scenarios derived from the roots of Hermite polynomials will emulate a normally weighted regression fit, noting that the Hermite polynomials are orthogonal with respect to the Normal distribution.

5.3.17 We note that much of the discussion here relates to precise interpolation of polynomial replicating formulae and although the principles could equally be applied to alternative basis functions it would be unusual in practice. For example, it is possible to calibrate a replicating portfolio of 100 market instruments using only 100 calibration scenarios but unlike polynomials there

is no quick and easy way of picking the best 100 scenarios. As such, a regression would usually be employed.

5.3.18 The impact of using precise interpolation compared to regression is considered in greater detail in our case study for replicating polynomials in section 6.2.

Optimal Components or Optimised Whole

5.3.19 A further choice to be made is whether to construct a model as the sum of components which have each been independently optimised or optimise the model after the components have been summed.

5.3.20 This choice is often already being made in many proxy models, sometimes unconsciously, and potentially without full consideration of the implications and impact on results. The most obvious example is where a different formula is constructed and then optimised for each product and then the total liability is the sum of each individually optimised formula. The reason for this approach is often to add a layer of granularity to the results, allowing better analysis of each product whilst also providing greater accuracy at the product level.

5.3.21 However, this accuracy at the product level may be at the expense of accuracy at the total liability level due to errors accumulating across products. Alternatively, all the product formulae could have been summed and then optimised in one go, thus targeting a minimum error distribution at the total liability level. However this would be at the expense of accuracy at the product level.

5.3.22 Applying the same principle to each formula, we see that the formula itself can be broken down into components and the same decision is required as to whether those components are optimised or the formula is optimised.

5.3.23 In the case of replicating polynomials, there arises a natural splitting of the formula into individual risk components plus a non-linearity component. This is particularly useful as much of the analysis performed in a risk management framework relates to individual risks (risk limits, scenario testing, what-if analyses, reverse stress testing etc.). As such, the choice over optimising components or the whole formula is particularly relevant to polynomial formulae.

5.3.24 In the same way that optimising a total liability formula will result in a sub-optimal fit to individual products, so optimising a product formula will result in a sub-optimal fit to the components of that product. In the case of a replicating polynomial this would be a sub-optimal fit for each individual risk function, potentially making the formula unsuitable for assessing the impact of individual risks. Likewise, a product formula constructed from optimised components will not provide as good a fit at product level as a product formula that has been optimised in one go.

5.3.25 The impact of this choice is considered in the case study for replicating polynomials in section 6.2. Additionally, a sample proof and analysis for a

two-factor 3rd order polynomial is offered in the appendix. From this example, some important conclusions are drawn

When optimising the whole formula:

- Adjusting the domain in one risk variable will change the fit of the marginal functions in other risk variables.
- The fit of marginal risk functions is impacted by any underlying non-linearity.
- The more severe the non-linearity between risk variables, the further from optimal the marginal approximation functions becomes.

When optimising components:

- Adjusting the domain in individual risk variables has no impact on the fit of the marginal functions in other risk variables.
- The fit of marginal risk functions are unaffected by the extent of any underlying non-linearity
- The more severe the non-linearity, the further from optimal the overall fit becomes.

5.3.26 The conclusion of this is that one must give careful consideration to the uses to which the proxy model will be put when deciding on the method of calibration. There is also the possibility of implementing different calibrations of the same model to derive the best results specific to different uses. However, one must ensure that results from different calibrations are made consistent with one another before communication to, and use by, management.

5.3.27 Note that, although a regression fit will usually be performed on the whole formula, the distinction here is not between interpolation and regression. The formula could just as easily be constructed by summing components that have each been optimised using regression.

5.3.28 This is to be contrasted with precise interpolation where, even though the whole formula will be calibrated in one go, the resulting formula will still be the sum of optimal components due to the selected fitting points being the same as when optimising components individually. In theory it is possible to solve for the fitting points that optimise the whole function but the number of formula terms make it an unrealistic proposition.

Summary of Options

5.3.29 In table 5.3.1 we summarise the options discussed here as applicable to some of the different formula types currently in use. Here we see that not all the options are available to all model types. For example, we have already seen that a replicating portfolio is usually associated with a regression fit being performed. This does not rule out precise interpolation being used though it is unlikely.

Table 5.3.1

Type of Proxy Formula	Determining Formula Structure	Regression, Interpolation or Both	Optimised Components, Whole or Both
Replicating Polynomials	Choice and number of nomials	Both Possible	Both Possible
Radial Basis Functions	Choice of radial basis function	Both Possible	Optimised Whole
Replicating Portfolios	Choice of assets	Regression	Optimised Whole
Commutation Functions	Choice and number of commutators	Both Possible	Optimised Whole

5.3.30 Another example is Radial basis functions which are usually optimised across the whole risk distribution rather than being built up from a series of optimised components.

5.3.31 Each of these model types are discussed in more detail in section 6 where we also addressing some of the issues discussed in this section.

5.4 A SPECIAL CASE – LEAST SQUARES MONTE-CARLO

5.4.1 Least Squares Monte Carlo (LSMC) has its roots in techniques for valuing American options where early exercise is possible, the technique originally being developed by Longstaff & Schwartz (2001).

5.4.2 LSMC is usually implemented as a polynomial replicating formula, although it isn't the use of a polynomial that characterises LSMC. The method could just as equally be applied to calibrate a replicating portfolio or any formula constructed from suitable basis functions. In this section, we give an overview of what sets LSMC apart from the other methods.

5.4.3 LSMC is not a different type of model but is characterised by the method it uses to calibrate the formula coefficients. The insight for LSMC is that under alternative methods, a lot of computational power is used to value a few heavy lift valuations leaving no computational power for other scenarios. Least-Squares Monte Carlo comes at the problem from a different angle attempting to recapture that power from the few specific valuations and to redistribute that computing power over more scenario valuations, each of which is less accurate. For example, instead of evaluating five scenarios accurately, each requiring two thousand stochastic simulations, the same computational power can be utilised to evaluate five thousand scenarios each using only two stochastic simulations.

5.4.4 Using only two simulations for each scenario will inevitably lead to large simulation errors across individual scenario results. However, the beauty of the method is that by performing a least squares regression fit against these inaccurate scenario results, the simulation errors across the fitting points tend to offset one another, and the resulting curve. The net error between the fitted

curve and the true curve is thus minimised without having to estimate the true curve or points on it.

5.4.5   LSMC is often implemented by optimising the overall fit rather than optimising components so there is the possibility that formula components may be sub-optimal. However, should the priority be to achieve greatest accuracy within formula components there is no reason why the method could not be adapted to optimise components.

5.4.6   Also, the method will often be implemented using automated algorithms to determine the optimal formula elements and coefficients at the same time. As discussed in section 5.2 this may give rise to issues around intuition[3]. It may also lead to unstable formula structures that vary from one data set to another or from one period to the next without providing insights into why the formula structure has changed. However, we again find that there is no reason why the method could not be implemented to determine formula coefficients only, leaving the user to determine the formula structure as they see fit. That said, despite potential challenges in interpreting the calibration, the considerations when validating the model are no different from any other calibration technique, requiring an understanding of the quality of the fit and the impact any limitations on the metrics which the model is used to produce

5.4.7   Finally, the large number of calibration scenarios used by LSMC requires generation through some method such as low discrepancy sequences (Glasserman, 2003). For each scenario, there needs to be a robust method for choosing the economic simulations for each real-world scenario. This includes re-scaling the economic simulation so it is applicable in the scenario chosen.

---

[3] Any of the methods can be set up in this way so this is not an issue exclusive to LSMC. Attention is drawn to the fact here due to LSMC usually being implemented through automated algorithms.

6 SPECIFIC MODELS DISCUSSED IN DETAIL

6.1 METHODOLOGY

6.1.1 Our aim here is to compare, across thousands of scenarios, the 'actual' value of some liability to those approximated by various proxy models. One of the problems associated with any proxy model is how to reliably assess accuracy without testing every single scenario. If we could run every scenario then a proxy model would not be required in the first place. In fact we can do exactly that by changing our point of reference from a stochastic cashflow model to something less complex. We can then evaluate every simulation result in our less complex model of reality and then re-evaluate in the various proxies. Analysis of the proxies can be performed and conclusions drawn which are no less valid than if a stochastic cashflow model had been used since the proxy models remain within definition, i.e. a model approximating a more complex model.

6.1.2 The model of reality used to produce 'actual' values was a purpose built cashflow projection model of a simple with-profit bond offering a maturity guarantee. The biggest simplification came from modelling time value using a Black-Scholes closed form solution rather than running thousands of simulations. In many other ways, the model retained the various complexities one might associate with a bond model used in a real life office environment such as guarantees increasing with regular bonuses and decrements from lapses, deaths and PUPs. The asset mix was also varied with term to maturity switching from equity toward fixed interest as maturity approaches. In this way the volatility used by the Black-Scholes formula varied across model points and was not a fixed value.

6.1.3 Over a thousand model points of varying term to maturity and 'moneyness' of guarantees were chosen. The asset share and cost of guarantees were evaluated separately for each model point, the total liability being the sum of total asset share and cost of guarantees across all model points. For simplicity we assumed that the cost of guarantees is backed by a fixed cash amount and the asset share liability matched so that capital results are derived purely from the variation in cost of guarantees.

6.1.4 Nine market and insurance risks were incorporated into the model as instantaneous time zero stress parameters in order to replicate common life office methodology whereby the one year Value-at-Risk is estimated from many instantaneous time zero stressed scenario results.

6.1.5 The modelled market risks are parallel yield shifts, UK equities, overseas equities, property, credit spreads and inflation whilst the modelled insurance risks are, mortality, persistency and expenses.

6.1.6 For simplicity, all of the risks are Normally distributed, expressed as stress percentages about a mean of zero with standard deviations given in table 6.1.1.

Table 6.1.1 Risk Parameterisation

Risk Parameter	Standard Deviation
Persistency	20.0%
Mortality	5.0%
Expenses	5.0%
Yield	0.75%
UK Equities	15.0%
Overseas Equities	17.5%
Property	7.5%
Credit Spreads	5.0%
Inflation	0.75%

6.1.7   In sample fitting points are drawn from both normal and uniform distributions whilst out of sample test points are drawn from the normal distribution only this being the distribution underlying each of the risks. Where uniformly distributed scenarios are used they are drawn from a domain between plus and minus four standard deviations, e.g. ±80% for persistency.

6.1.8   The unstressed base liabilities are as follows:

Asset Share (£m)	Cost of Guarantees (£m)	Total Liability (£m)
1,728	232	1,959

6.1. 8   Replicating polynomials would seem to offer the most accessible means of demonstrating each of the concepts discussed in earlier sections so it is this type of proxy model we consider first as a case study before performing similar analyses for other model types.

6.2   REPLICATING POLYNOMIALS

Introduction

6.2.1   Within the UK insurance space there has been a significant increase in the use of replicating polynomials in recent years. This is due in part to their incorporation within mainstream software solutions available to industry. However the uptake by insurers and developers alike has most likely been driven by the speed and apparent simplicity of using a polynomial to reproduce complex liability values.

6.2.2   Polynomials are indeed very quick to evaluate and theory tells us that a replicating polynomial can be constructed to any degree of accuracy (Stone, 1948). As a result, the lure of being able to run tens or even hundreds of thousands of scenarios has been a tempting proposition. Unfortunately, the reality is not quite so simple and the practical issues to resolve are numerous.

6.2.3   As well as being simple and quick to calculate, polynomials are relatively simple to understand as a functional form. However, this simplicity quickly disappears as the number of risk variables increase. In the language of our replicating formulae construction, the basis functions, X_k, are very simple functions of the risk variables, r_n.

6.2.4 As discussed in section 3.3, the simplicity of each formula element means that a large number of elements are required in order to model complex behaviours. For polynomial replicating formulae, this is manifested through an increase in the order of the polynomial. Also, as the number of risk variables increase so the potential number of formula elements increases exponentially. As the number of formula elements increases, so too does the minimum number of heavy model runs that are required to determine the coefficients, β_k, of each element. Very quickly, the light model may not seem so light any more.

Determining Formula Structure

6.2.5 With the potential for many hundreds, or even thousands, of combinations of formula elements, the use of an automated algorithm to test every possible formula structure would appear to be an obvious solution. However, as discussed in 5.2, an automated algorithm will often be simply an emulation of expert judgement through the codification of subjective decision making.

6.2.6 When it comes to determining formula structure, there are a huge number of ways to proceed and the potential for different decision making processes is large. Our interest here is the decisions that need to be made and not the codification of those decisions. As such, we proceed with the determination of formula structure in a completely subjective way in order to look more closely at one possible decision making and design process.

Marginal Risk Functions
6.2.7 A replicating polynomial can easily be broken down into its component parts, consisting of a number of univariate polynomials combined with a number of multivariate polynomials. Each univariate polynomial, referred to as a marginal risk function, represents the variation in value with respect to a single risk and it is these marginal risk functions that we address first in determining our replicating polynomial formula.

6.2.8 We start by considering the variation in the value of cost of guarantees (CoG) with respect to persistency risk. Asset share is not considered at this stage as it does not vary with insurance risks. We begin by trying a quadratic proxy first, for efficiency we use precise interpolation rather than regression as this requires only three fitting points or nodes. We can base the fitting points on the roots of the Legendre polynomials which will provide the best estimate of the fitting points for the optimal least squares solution (Hursey & Scott, 2012).

6.2.9 The chart in figure 6.2.1 shows the variation in cost of guarantee with respect to persistency risk for both the primary and proxy models. The resulting error curve is shown in figure 6.2.2.

6.2.10 The error curve is very close to the third order Legendre curve demonstrating that it is close to an optimal fit for a quadratic polynomial proxy as shown by Li, Y,M (n.d.). The implication of this is that to improve the fit further would require a higher order polynomial.

Fig. 6.2.1 Persistency Risk, Actual vs Estimated
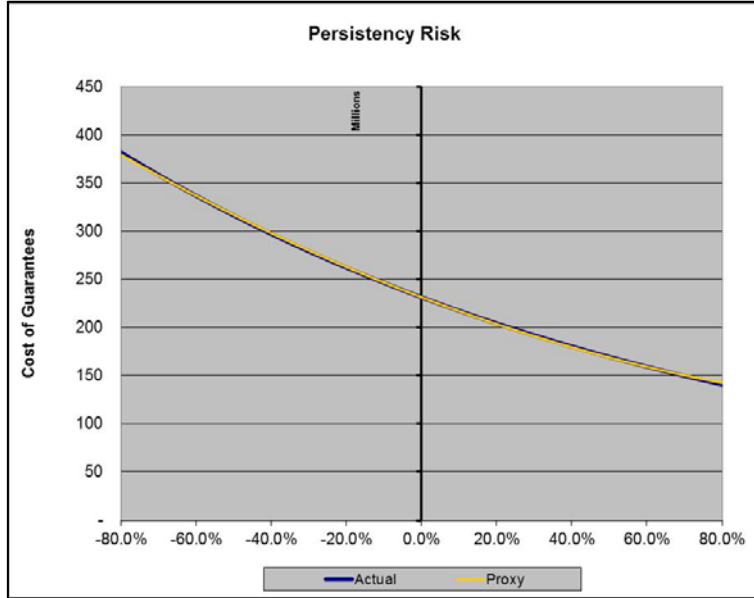
Fig 6.2.2 Persistency Risk, Error Curve
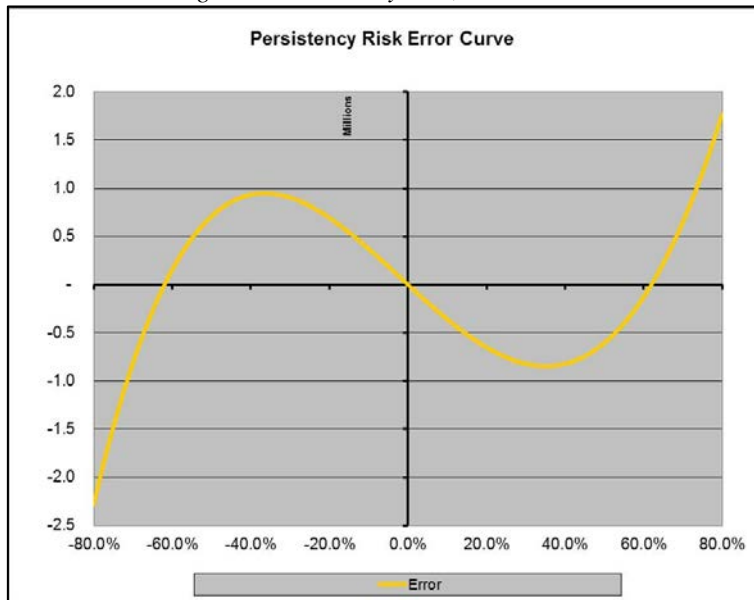
6.2.11 Based on this error curve we can either deem a quadratic to be a sufficient order of polynomial for modelling persistency risk or demand a closer fit through a higher order polynomial, the fit already being optimal for a quadratic polynomial. This is further demonstrated in figure 6.2.3 where we show that a regression fit using 100 calibration scenarios drawn from a uniform distribution provides the same fit as precise interpolation using Legendre polynomials.

6.2.12 We also consider here the impact of performing a regression fit if the calibration scenarios are drawn from a Normal distribution. Figure 6.2.4 shows the error curves that result from calibration scenarios drawn from the Normal and Uniform distributions.

Fig. 6.2.3 Precise Interpolation vs Regression
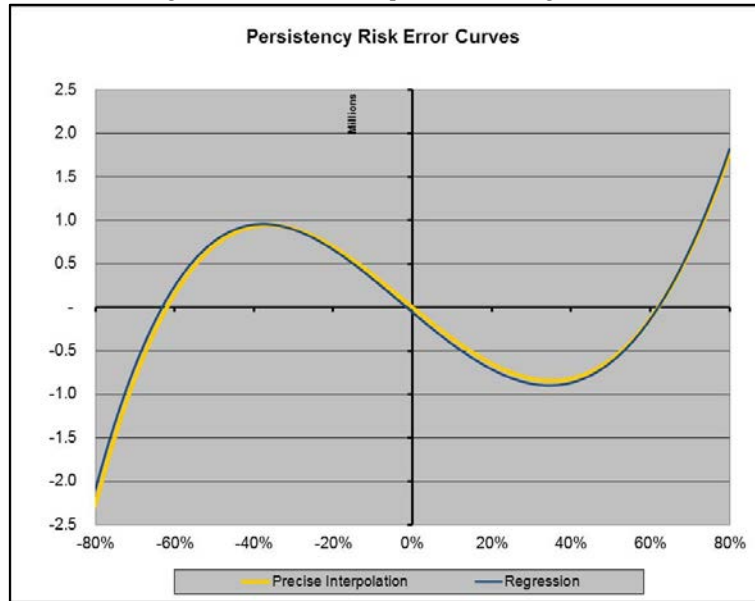
Fig 6.2.4 Normal vs Uniform calibration scenarios
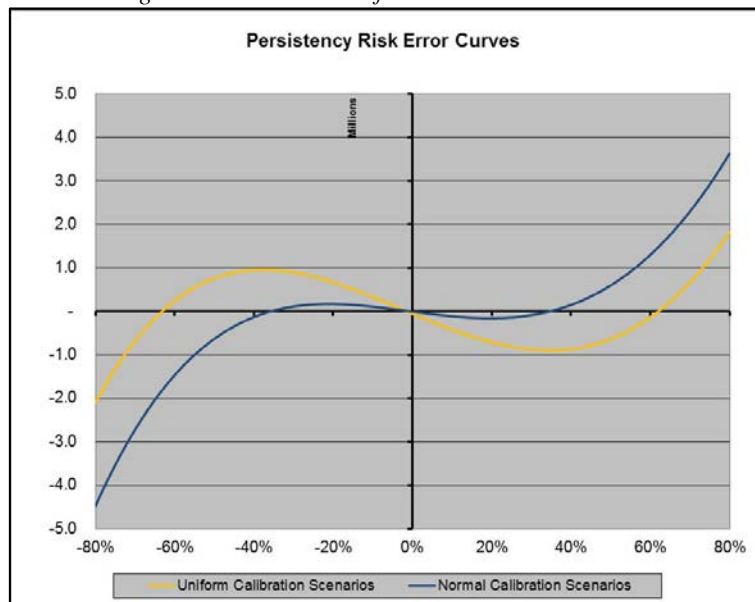
6.2.13 The greater mass of points in the centre of the uniform distribution lends more weight in this area thus providing greater accuracy at the centre but at the expense of larger errors at the tail of the distribution. In fact, drawing calibration scenarios from a normal distribution and then performing a least squares fit is equivalent to performing a weighted least squares fit, the weights provided by the normal distribution.

6.2.14 We have already observed that the roots of the error curve that result from Uniformly distributed calibration scenarios correspond to the roots of the 3rd order Legendre polynomial. It can also be shown that the roots of the error curve resulting from Normally distributed calibration scenarios correspond to the roots of the 3rd order Hermite polynomial.

6.2.15 In this example, the 3rd order Hermite polynomial predicts roots of the error curve at 0 and ±34.64% (risk standard deviation of 20% multiplied by ⎯). The actual roots occur at -35.7%, 0, and +34.69%. This demonstrates that the regression fit could have been reproduced through precise interpolation using the Hermite roots. A proof is provided in the appendix.

6.2.16 All this serves to demonstrate that care must be taken with the calibration scenarios when employing a regression fit since points drawn from anything other than a Uniform distribution will provide a sub-optimal fit in the traditional least squares sense, i.e. unweighted.

6.2.17 Returning to our original quadratic fit in figure 6.2.2, we decide in this case that the fit is sufficient for our purposes, noting that the maximum error at £2.3m is less than 1.0% of base CoG.

6.2.18 Similar exercises are carried out for the other insurance risks; mortality and expenses. Precise interpolation is used throughout with the fitting points again being determined using the three roots of the third order Legendre polynomials. The resulting error curves are shown in figures 6.2.5 and 6.2.6.

Fig 6.2.5 – Expense Risk	*Fig 6.2.6 – Mortality Risk*
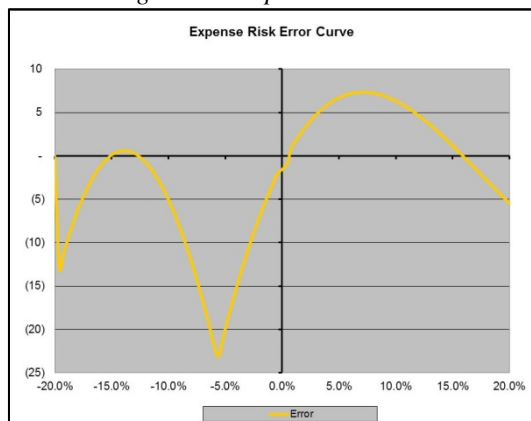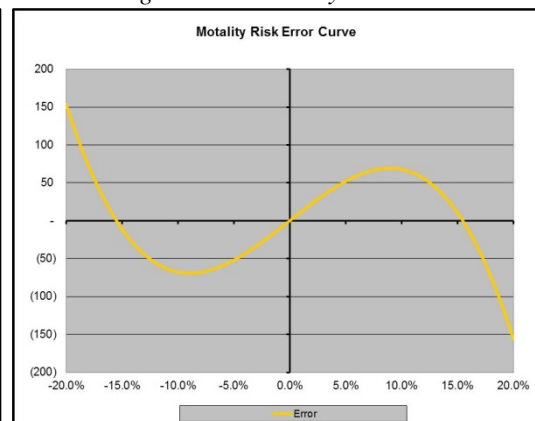
6.2.19 Errors are less than £200 for mortality risk and less than £25 for expense risk compared to a variation in CoG with respect to each risk of £11m and £6m respectively. Such a close fit despite the movement in CoG is perhaps indicative of insufficient complexity in the cashflow model used to generate results but in the author's opinion it does not invalidate the analysis. However, such small errors lead us, again, to deem a quadratic proxy sufficient for modelling each of these risks.

6.2.20 We get similar results when we test various market risks. UK equity risk, overseas equity risk, and spread risk are shown in figure 6.2.7 along with the error curves resulting from fitting quadratic polynomials.

6.2.21 Whilst the error curves for overseas equity and spread risk indicate an optimal fit for a quadratic, the error curve for UK equity risk indicates that the fit is not optimal and could be improved further without an increase in polynomial order. However, even with a sub-optimal fit, the maximum error for UK equities is less than £600k. Maximum errors for property risk and inflation

risk (not shown here) are significantly less. Therefore a quadratic is, again, deemed sufficient to model each of these market risks.

Fig. 6.2.7 – Market Risks
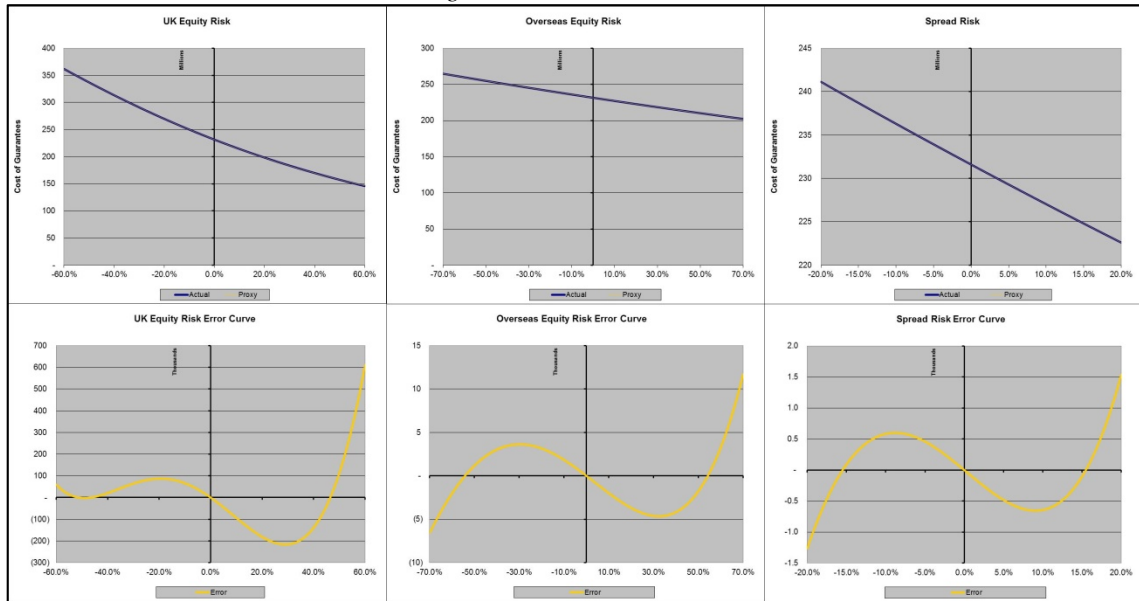
6.2.22 The analysis so far relates only to the cost of guarantee liability. Asset share is also sensitive to market risks but due to this sensitivity being linear for all but the interest rate risk, the proxy marginal risk functions trivially fit the data perfectly in respect of these risks. Consequently, no further analysis is offered.

6.2.23 Things gets a little more interesting when we consider the fit for interest rates. We begin with CoG, showing the variation with respect to parallel yield shifts in figure 6.2.8. The behaviour of the curve at low interest rates leads to a quadratic not providing an ideal fit as it fails to capture the flattening out of the curve due to the flooring of interest rates.

Fig 6.2.8 – Interest Rate Risk, Actual vs Estimated
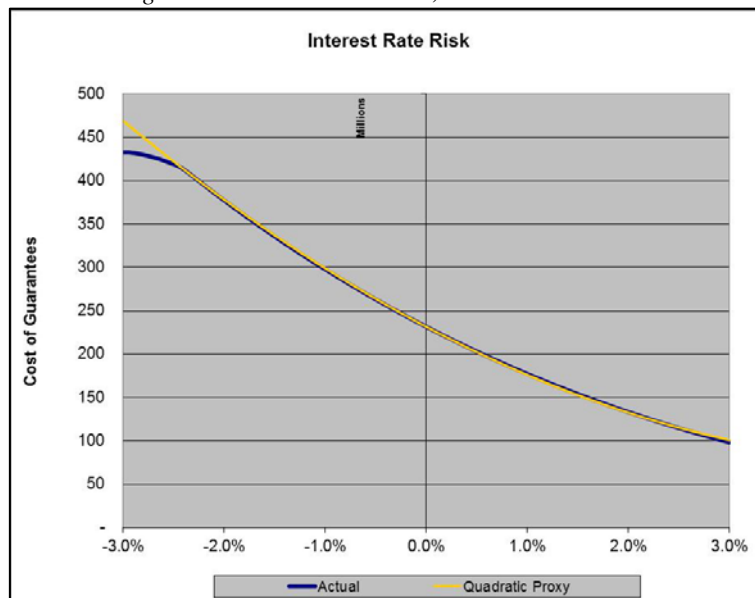
6.2.24 As a result, we tested higher order polynomials – cubic and quartic – to see if the fit can be improved. The resulting error curves are exhibited in figure 6.2.9. Unfortunately, even with higher order polynomials, the fit remains unsatisfactory at the lower end of the domain.

6.2.25 As we have been using precise interpolation to fit and assess the formulae, we investigate the possibility that it is the use of prescribed fitting points that leads to a poor fit at the extreme low end of the curve. Potentially, performing a regression fit using a larger number of points drawn from the whole domain may capture the shape of the whole curve more effectively and improve the fit. We again test quadratic, cubic and quartic polynomials. The chart in figure 6.2.10 shows the error curves resulting from regression fitting.

Fig. 6.2.9 – Error Curves, Interpolation	*Fig. 6.2.10 – Error Curves, Regression*
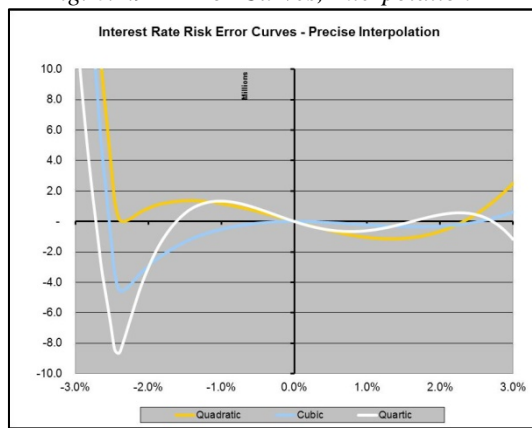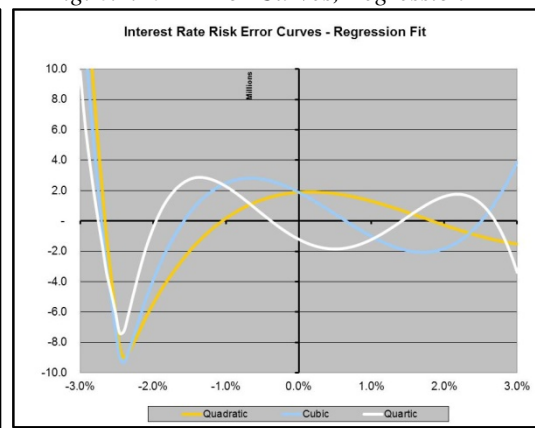
6.2.26 Comparing the regression fit alongside the interpolation fit, some improvement is observed at the lower end of the curve but at the expense of the fit elsewhere. However, it would still appear that a much higher order polynomial or another type of function may be required to capture more precisely the behaviour of the curve at low interest rates.

6.2.27 The decision over how to proceed from here is not straightforward. An increase in order of the polynomial to a quartic may be justified on the grounds that the error is much lower at the very tail of the distribution.[4] However, if we were to set a materiality limit of, say, £5m then based on the interpolation results it could be argued that a cubic or quadratic should be favoured over the quartic as materiality is not breached until interest rates fall by more than 2.5% compared to 2% for the quartic. A lower order polynomial also has the advantage of reducing complexity by lowering the potential number of cross terms later in the design process. If we consider the regression results, the justification for a move to a quartic is even less clear with all three error curves exhibiting very similar traits.

6.2.28 Note here the level of subjective judgement that is required. Clearly, metrics could be drawn and a more objective decision made and this would have the advantage of being repeatable and perhaps codified and automated. However,

---

[4] An alternative is to split the domain and fit to each sub-domain separately. There are disadvantages, however, not least in the potential increase in number of fitting and test scenarios.

the subjective analysis provided here shows us that a more sophisticated measure than least squares would be required if we wished to do so. In fact, we have here an example of less weight being given to the outcomes at the lower tail (ref. 5.3) of the distribution due to other considerations. Applying a mechanical weighted least squares process to a subjectively designed weight function may therefore offer a potential route to automation.

6.2.29 In this example, we opt for a quadratic fit as we will be employing both interpolation and regression fits for analyses. We also note that our domain extends to four standard deviations. Based on this, the 99.5[th] percentile stress will be around 2% at which the quadratic should prove sufficient within a materiality limit of £5m. Also note here that we have also derived an approximate limit for the model, being aware that stresses beyond a fall of 2.5% will 'break' the model.

6.2.30 This example also serves to highlight the importance of communicating any limitations to management, along with a clear explanation of the consequences of those limitations. In the above example we have an approximate limit of 2.5% beyond which the model should not be relied upon to assess the impact of falls in yields. If the model is used to "roll forward" results in a decreasing interest rate environment then the biting point for failure will reduce even further. Additionally, once a model is in a production environment it will undoubtedly be called upon to produce analyses and metrics beyond that envisaged by those that initially set up the model; queries from regulators and other ad-hoc investigations may not be known in advance of calibration. If limitations are not advertised sufficiently then the model may unknowingly be used beyond its limits thus providing misleading or incorrect information.

6.2.31 Turning our attention to the asset share we note that the fit exhibits the same properties as the fit to cost of guarantees albeit with lower magnitude of errors. Figure 6.2.11 illustrates.

Fig. 6.2.11 – Asset Share Error Curves
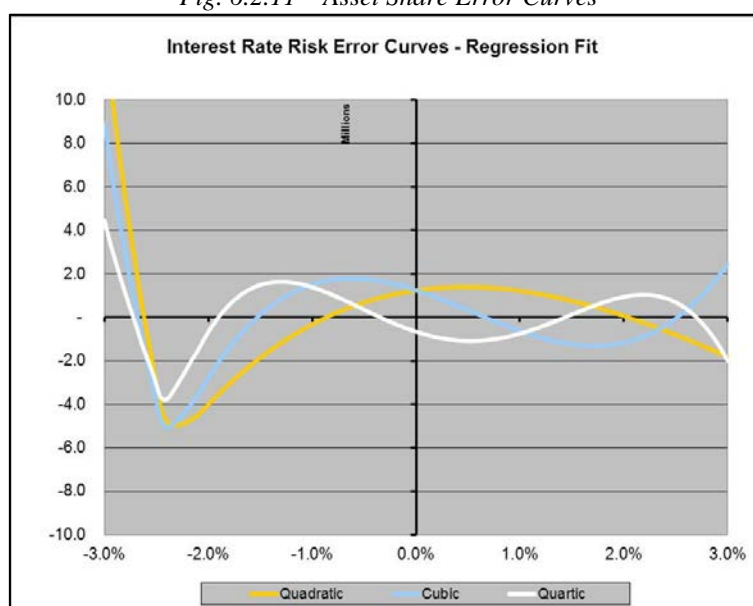
6.2.32 Noting that the errors are much smaller as a percentage of asset share and following the same reasoning as for the fit to cost of guarantees we decide a quadratic is sufficient for the purposes of analysis. Additionally, noting the lack of dependence on insurance risks and the trivial nature of the dependence on other market risks we limit our interest to the cost of guarantees for the remainder of the formula design process reasoning that a formula sufficient to model cost of guarantees should be sufficient for the much less complex asset share.

6.2.33 Ultimately, we have a quadratic marginal risk function for each of the nine risks which can be added together to build a replicating formula ignoring non-linearity effects. Non-linearity and the risk dependency structure are considered next.

Non-Linearity and Risk Dependency Structure
6.2.34 Having determined all the marginal risk function we turn our attention to non-linearity and the risk dependency structure.

6.2.35 Non linearity is the difference between the combined impact of two or more risk factors and the sum of the marginal impacts of those same factors. So if an equity scenario leads to a stress of 40 and a lapse scenario leads to a stress of 60 and the combination of both leads to a stress of 110 we have a non-linearity impact of 10. Given that the marginal risk functions deliver a value of 100, we require a function in both lapse and equities to deliver the additional 10. We refer to this multivariate function as the non-linearity surface.

6.2.36 To derive our first non-linearity surface we consider the risk pairing Persistency and UK equity. We begin by constructing a combined risk surface adding together the two marginal risk functions. The resulting surface is illustrated in figure 6.2.12. Deducting this from the actual combined risk surface (evaluated using our primary heavy model) allows us to evaluate the non-linearity surface, illustrated in figure 6.2.13.

Fig 6.2.12 Combined Risk Surface *Fig 6.2.13 Non-Linearity Surface*
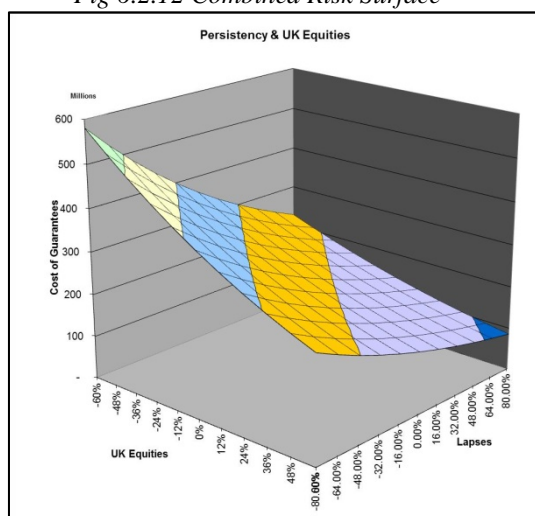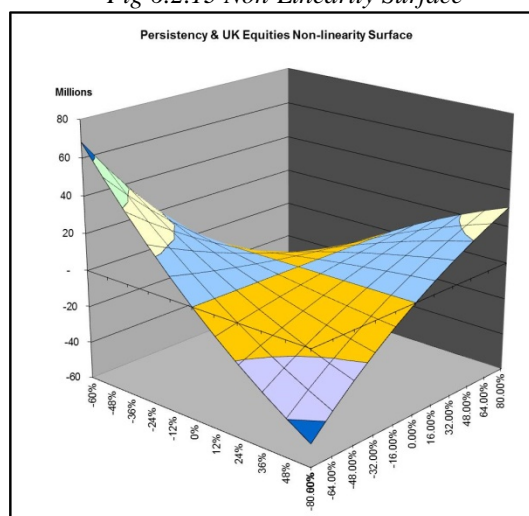
6.2.37 We can now attempt to construct a two factor polynomial approximation to non-linearity starting with a simple *xy* cross term. Deducting from our non-linearity surface leaves the error surface depicted in figure 6.2.14.
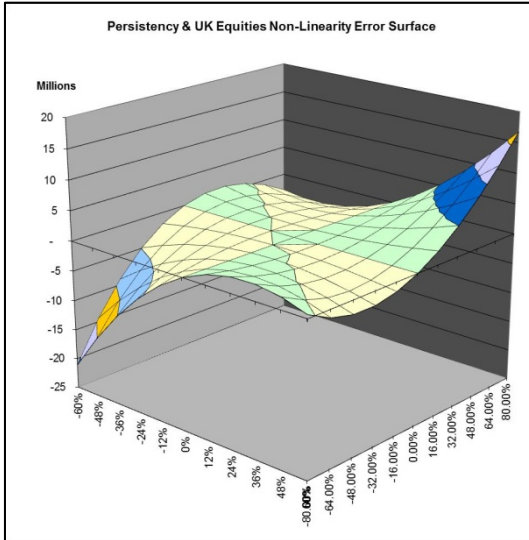
Fig. 6.2.14 – Non-Linearity Error Surface, xy term only

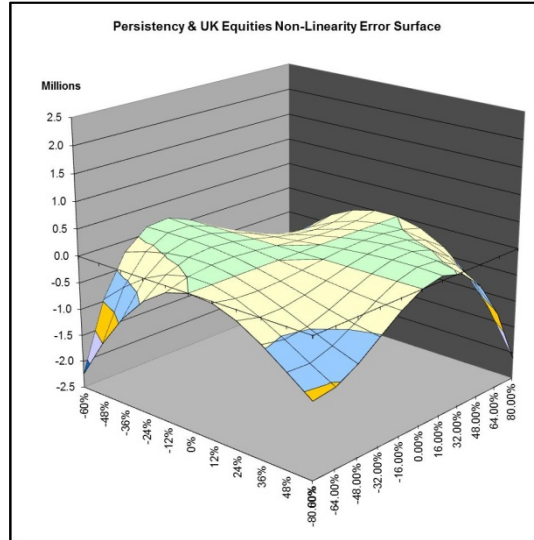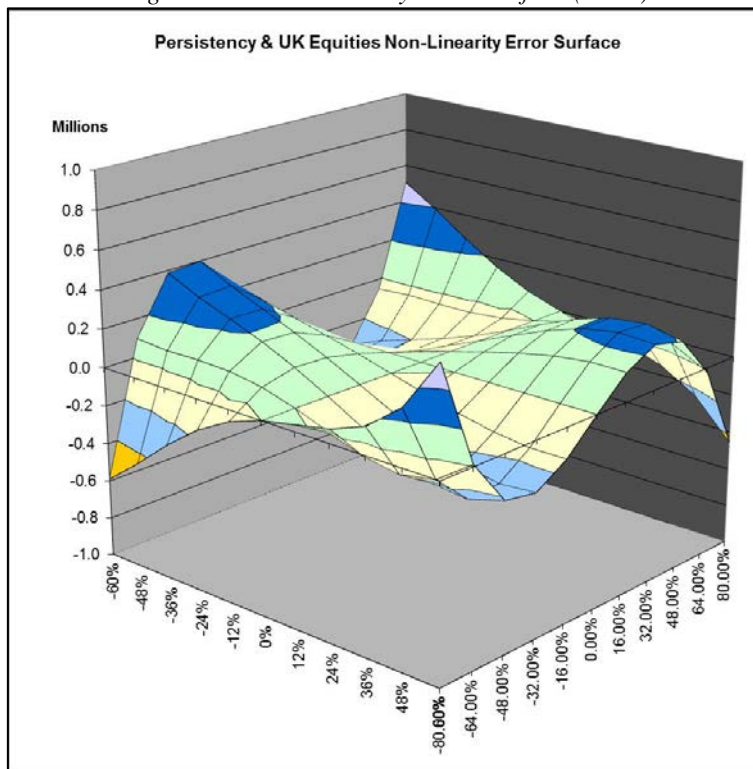*Fig. 6.2.15 – Non-Linearity Error Surface, xy, x^2y and xy^2 terms*

Fig. 6.2.16 – Non-linearity Error Surface (Final)

6.2.38 The problem with using a single *xy* cross term is that it is a symmetrical and linear function in both *x* and *y*. Non-linearity is rarely either so a single *xy* cross term will often provide a poor fit to non-linearity. By using a combination of terms in *xy*, x^2y and xy^2, the fit can be improved significantly as shown by the error surface in figure 6.2.15 (note the change in axis scale). The fit can further be improved with the addition of a fourth x^2y^2 term as

illustrated in figure 6.2.16, where the magnitude of the errors are reduced below £1m across the whole surface.

6.2.39 Thus the formula component for non-linearity between persistency risk and UK equity risk will be of the form:

$$c_1xy + c_2xy^2 + c_3x^2y + c_4x^2y^2$$

where x and y are the persistency and UK equity stresses and $c_1,...,c_4$ are constants.

6.2.40 Just as we did for the marginal risk functions in one dimension, our two dimensional non-linearity functions can potentially be calibrated using either regression or precise interpolation. We demonstrate this when considering the non-linearity between persistency and interest rates which ultimately proves to have the largest non-linearity impact ranging from +£150m to -£100m as shown by the non-linearity surface in figure 6.2.17.

Fig. 6.2.17 – Non-Linearity Surface
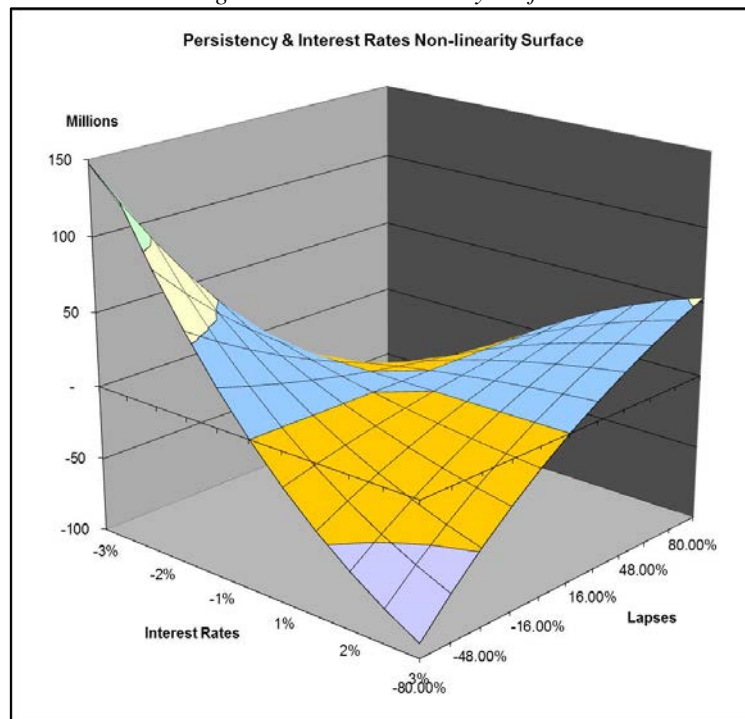
6.2.41 A non-linearity function of the same form as that derived for persistency and UK equity risk is fitted to this non-linearity surface, first by least squares regression using one hundred fitting points and then using precise interpolation. The four fitting points for interpolation are derived from the intersection of the non-zero roots of the 3rd order Legendre polynomials for each of the two risk parameters. The resulting error surfaces are shown in figures 6.2.18 and 6.2.19 respectively.

6.2.42 Using a least squares regression fit provides a better overall fit than when precise interpolation is used, as measured by sum squared errors and maximum error. However, the nature of our interpolation fitting points leads

to a better fit for a large portion of the domain. In fact, it is only at one corner of the domain that the fit is worse than the regression fit. If this corner is excluded then the interpolation fit is better in terms of both maximum error and sum squared errors

Fig. 6.2.18 – Error Surface, Regression Fit   Fig. 6.2.19 – Error Surface, Interpolation
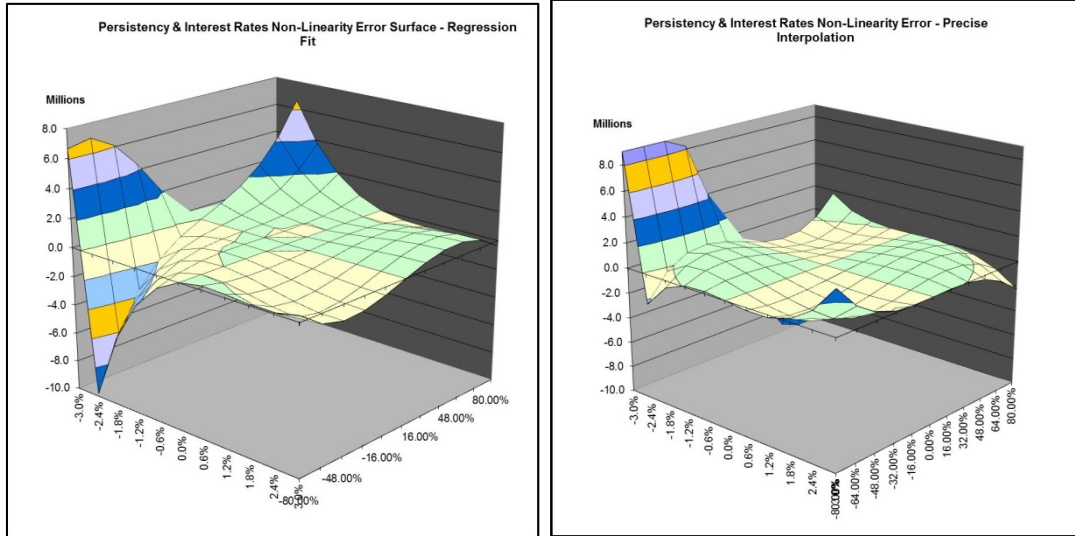
6.2.43 If we consider that the likelihood of events in this corner are a combination of two extreme events it may be that these events are outside our region of interest and the interpolated fit is preferred. Alternatively, very extreme events may be of particular interest in which case the regression fit would be preferred. Once again, the importance of considering the use to which the model is put comes to the fore. Note also that we have another example where a sub-optimal fit, in the least squares sense, may be chosen due to other subjective considerations. However, this decision over the preferred fit can be codified and potentially automated if we realise that the sub-optimal fit in the least squares sense can be chosen by attaching less weight to the corners of the surface through application of an appropriate weight function.

Completing the formula structure
6.2.44 The formula components for each of the risk pairings considered so far each have four coefficients to determine thus requiring at least four heavy lift calculations to calibration each one. Even with only nine risks, there are thirty six possible pairs to consider so the potential number of formula terms is already quite high before we even consider interactions between three or more risks. Process automation may help but a large number of scenario results would still be required in order to perform the analysis and determine formula structure effectively.

6.2.43 We must therefore proceed in a methodical manner, where possible, making use of any knowledge we have to eliminate risk interactions from the investigation. Also, where non-linearity effects are trivial they can be excluded from the formula structure.

6.2.44 Completing our analysis we find that risk interactions involving expenses, mortality and inflation could be ignored altogether as being either trivial or non-existent. For risk pairings involving persistency and interest rates, two factor polynomials of the form already described were used. All other risk pairings were also ignored, e.g. equities and property, leaving nine risk pairings to be modelled.

6.2.45 One further formula component was added, that being a three factor non-linearity function combining the three largest risks, persistency, interest rate and UK equities. The purpose is to capture any residual non-linearity arising from the three risks acting together, over and above that already captured by the non-linearity components derived from pairs of risks.

6.2.46 Putting it altogether we have a formula which consists of nine quadratic marginal risk functions, each consisting of two terms plus a single combined constant term for all nine. We then have nine two-factor non-linearity functions each consisting of four terms plus one further three factor non-linearity function consisting of eight terms. This gives a total of twenty formula components of sixty three terms. Table 6.2.1 summarises.

Table 6.2.1 – Replicating Polynomial Summary

Formula Component	Number of Components	Number of Elements
Constant	1	1
Quadratic Marginal Risk Function	9	18
2 Factor 2nd Order Non-Linearity Function	9	36
3 Factor 2nd Order Non-Linearity Function	1	8
TOTAL	20	63

Optimised components vs Optimised whole

6.2.48 Now that the formula structure has been determined we can calibrate and test the formula using various datasets. Before proceeding, however, we take a brief look at the impact of optimising the formula as a whole compared with building the formula from optimised components (ref. 5.3).

6.2.49 Whilst precise interpolation of the whole formula will continue to optimise components (due to the choice of fitting points), regression fitting the whole formula will optimise the whole formula and in doing so will lead to sub-optimal fit for formula components.

6.2.50 Recall the optimised persistency risk error curves in figure 6.2.3. These are the error curves resulting from the optimisation of the formula component for persistency risk, one by interpolation and the other through regression, but both leading to the same result. The chart in figure 6.2.20 compares that curve with the marginal risk error curve that results when the whole formula is optimised together in one go.

6.2.51 As can be observed, optimising the whole formula has led to a loss in quality of fit of this formula component, and potentially others. In particular, the error at the tail of the distribution is significantly worse than when the formula components were optimised individually.

*Fig. 6.2.20 – Optimised Components vs Optimised Whole*
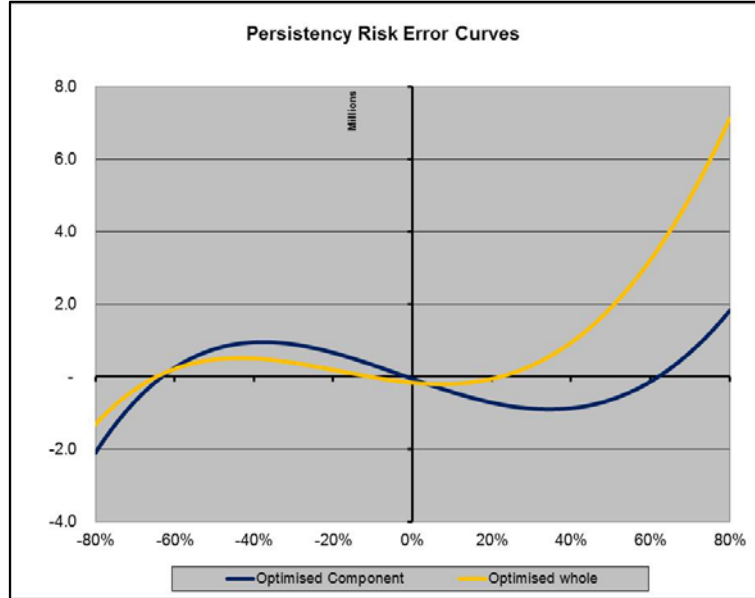
6.2.52 If, say, we wished to use this model to measure individual risks and risk capital components, then optimising those components will provide better answers, noting here that both of the curves in figure 6.2.20 were derived using regression. Once again, use of the model must be consideration when deciding the method of calibration.

Results

6.2.53 We now consider the results produced by a number of different calibrations. We consider both precise interpolation and regression using between 100 and 1,000 fitting points. 4,500 out-of-sample test points are used to calculate the asset share and cost of guarantee liabilities.

Fig. 6.2.21 – Asset Share, Actual vs Estimated
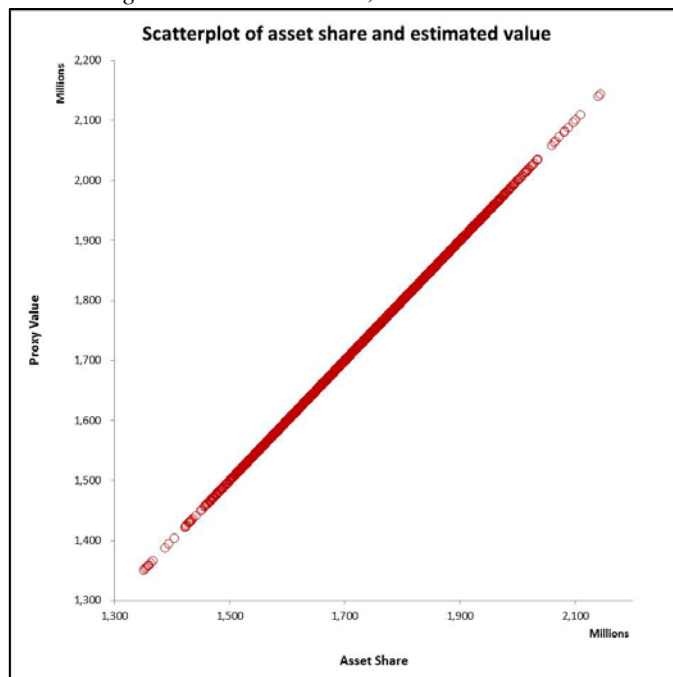
6.2.54 We start by testing the proxy for asset share. The chart in figure 6.2.21 shows the true asset share for the out of sample points plotted against the estimated value when 100 fitting points were used.

6.2.55 It is immediately obvious that the fit is very good. The goodness of fit is confirmed by the chart in figure 6.2.22 which plots the percentage error in the approximation against the true value of the asset share.

Fig. 6.2.22 – Asset Share Error%
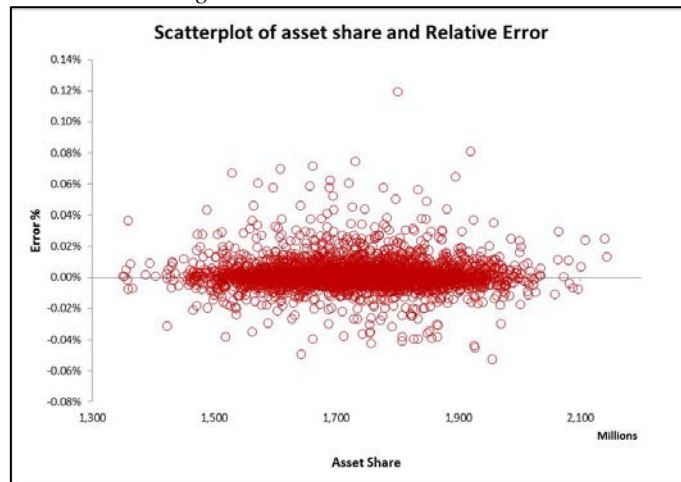
6.2.56 The maximum error is only 0.12% and the root mean square error is 0.01%. If we consider the linear nature of the market stresses, it is perhaps to be expected that a good fit would be achieved using a polynomial proxy. We now turn our attention to cost of guarantees.

6.2.23 – CoG, Actual vs Estimated
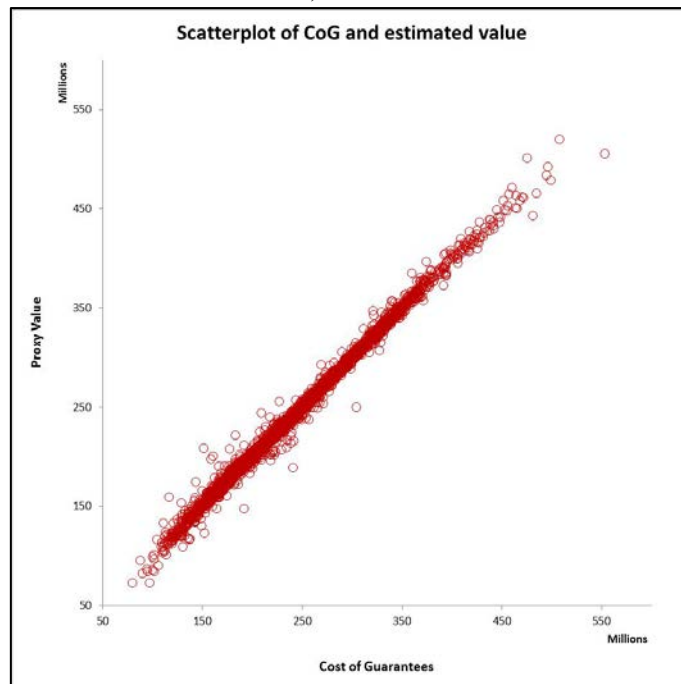
6.2.57 Asset share is assumed matched and VaR driven by variation in CoG so given the more complex (non-linear) and more interesting behaviour of CoG we restrict our analysis to CoG for the remainder of this section.[5]

6.2.58 The chart in figure 6.2.23 shows the CoG for the out of sample points plotted against the estimated value when 100 fitting points were used. The percentage errors in the approximation are plotted in the chart in figure 6.2.24.

6.2.58 The maximum error magnitude is significant at over £50m (53%). Root mean squared error is £4.6m (2.45%). The fit could be considered inadequate for the purposes of risk management. However, given that a 63 term formula is being calibrated using only 100 fitting points it is perhaps not surprising.

Fig. 6.2.24 – Cost of Guarantees Error %
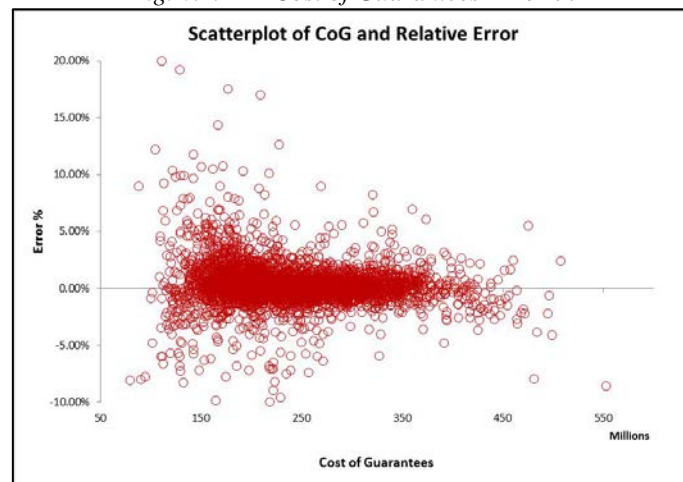
6.2.59 We now investigate the impact of increasing the number of fitting points, still fitting by regression, or reducing the number of fitting points and fitting by interpolation. For the regression fit we increased the number of fitting points to 400. For the interpolation fit we reduce the number of fitting points to 63, the minimum possible for a unique solution. The interpolation result is very sensitive to the fitting points used so they are selected based on the roots of Legendre polynomials. See Hursey & Scott (2012) for theoretical justification and explanation.

6.2.60 The charts in figure 6.2.25 show the scatter plots of actual versus estimated for each of the two calibrations along with the corresponding scatter plots of relative error.

6.2.61 Both calibrations have significantly improved the fit over that using 100 fitting points. The overall quality of fit appears very similar between the two although precise interpolation does, in this case, appear to give a slightly better fit than regression using 400 fitting points. This observation is confirmed if we compare maximum error, £22m (14.3%) versus £15m (7.2%), and root

---

[5] We did consider a combined fit to total liability (Asset share + CoG) but found that the errors in CoG were second order compared to the size of asset share, masking valuable insights and analysis.

mean squared error, £1.8m (0.9%) vs £1.7m (0.8%), for regression and interpolation respectively.

*Fig. 6.2.25 – Cost of Guarantees, Interpolation vs Regression (400 fitting points)*
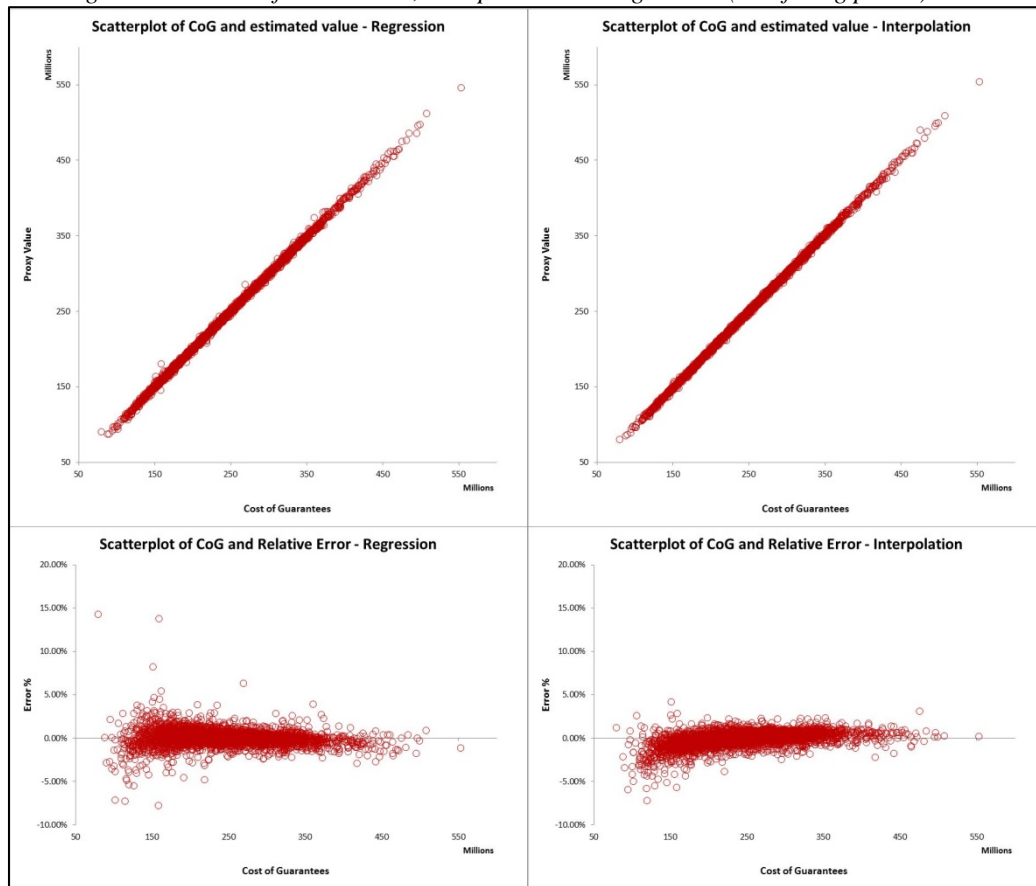
6.2.62 Further regression calibrations were tested ranging from 200 to 1000 fitting points. The results of a number of metrics were measured and are given in table 6.2.2 for each of the regression calibrations and the interpolation.

Table 6.2.2

No. of Calibration Scenarios	100	200	300	400	500	750	1000	63
Average Absolute Error (£m)	2.5	1.5	1.3	1.2	1.1	1.0	1.0	1.3
Root Mean Squared Error (£m)	4.6	3.3	2.2	1.8	1.7	1.6	1.5	1.7
Min Error (£m)	(53.4)	(20.5)	(12.0)	(12.3)	(12.9)	(13.9)	(11.8)	(9.1)
Max Error (£m)	57.8	83.5	34.5	21.8	19.6	9.8	9.0	14.8
Average Absolute % Error	1.2%	0.7%	0.6%	0.5%	0.5%	0.5%	0.5%	0.6%
Root Mean Squared % Error	2.5%	2.6%	1.3%	0.9%	0.8%	0.7%	0.7%	0.8%
Min % Error	-53.3%	-19.4%	-11.4%	-7.8%	-8.2%	-8.8%	-7.5%	-7.2%
Max % Error	38.3%	102.7%	35.6%	14.3%	13.5%	6.6%	5.7%	4.2%
R-squared	99.73%	99.87%	99.94%	99.96%	99.96%	99.97%	99.97%	99.97%

6.2.63 The charts in figure 6.2.26 plot various metrics from table 6.2.2 against the number of calibration scenarios to better illustrate and compare the quality of fit of each calibration. The interpolation fit is also plotted for comparison.

Fig. 6.2.26 – Goodness of fit by number of Fitting Points
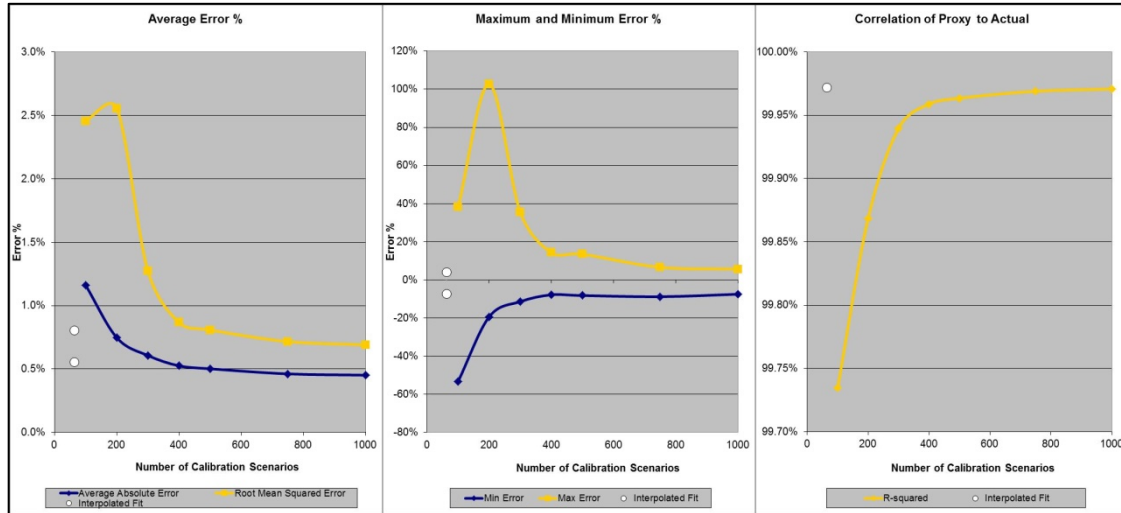
6.2.63 For most metrics the quality of regression fit improves as the number of fitting points increases. However, the law of diminishing returns applies with the rate if improvement decreasing as the number of fitting points increases. We also found that the interpolation fit gave near optimal results, being of an equivalent quality of fit to that achieved using between four and five hundred fitting points under regression fitting.

6.2.64 We now turn our attention to the 1-in-200 VaR as measured using each of the proxy calibrations. The VaR estimates and their error from actual of £445.8m are shown in table 6.2.3.

Table 6.2.3

No. of Calibration Scenarios	100	200	300	400	500	750	1000	63
1-in-200 VaR	440.8	441.4	443.3	442.4	442.4	443.3	442.9	444.5
Error	(5.0)	(4.4)	(2.4)	(3.4)	(3.4)	(2.5)	(2.9)	(1.3)
Error %	1.1%	1.0%	0.6%	0.8%	0.8%	0.6%	0.7%	0.3%

6.2.65 The errors at the extreme of the distribution are smaller than perhaps expected given the other metrics in table 6.2.2. Also, the correlation between VaR error and number of fitting scenarios is not as strong as for other metrics.

6.2.66 According to most conventional metrics, the regression calibration using 100 points is a poor fit as shown in table 6.2.2. However, the capital estimation error at the tail of the distribution is only £5m (1.1%). This is illustrated in the chart in figure 6.2.27 which shows ranked actual versus ranked approximations for this calibration. It shows that despite the poor fit at a scenario level, the distribution of the proxy is very close to that of the actual. Taking various quantiles is effectively just comparing results at different points along these two lines which are a close fit.

Fig. 6.2.27- Ranked Actual vs Ranked Proxy
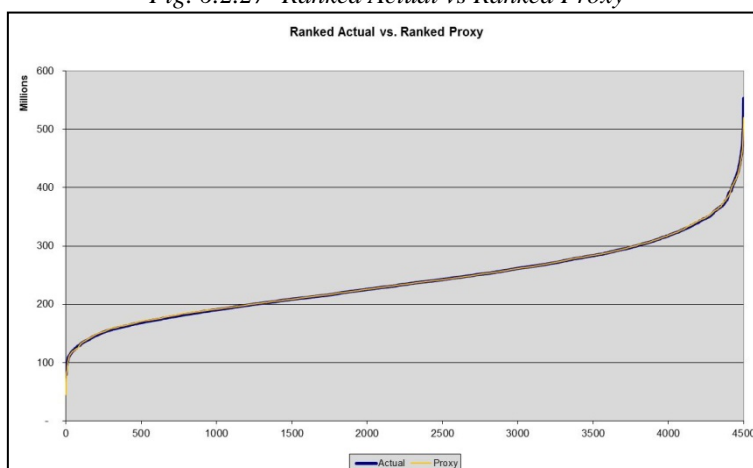
6.2.67 This can be illustrated more clearly on the chart in figure 6.2.28 where we plot the difference between the two lines from the chart in figure 6.2.27 (the difference between the ranked proxy results and the ranked actual results). Over this we have plotted the ranked errors (from figure 6.2.24). The ranked errors are as expected, representing the distribution of errors. The shape of the other line was not as expected, showing a good fit across nearly the whole distribution. This would lead to an accurate capital result despite the inherent inaccuracy of the model. An explanation is offered in section 4.2.

Fig. 6.2.28 – CoG Error, 100 point Regression
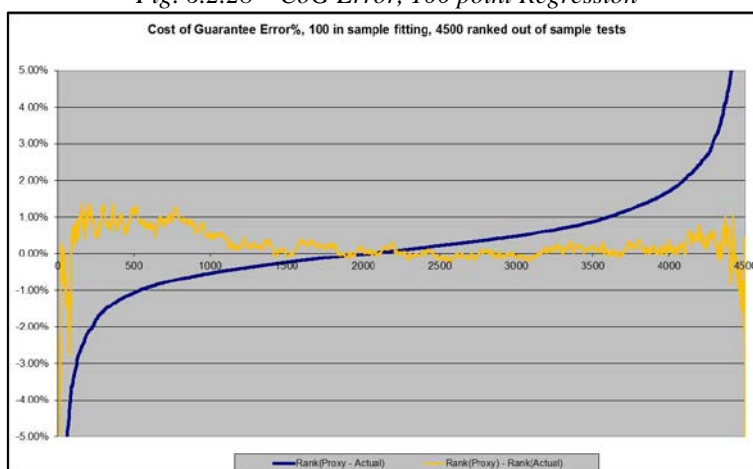
6.2.68 Plotting the same charts for other calibrations seems to show that increasing scenario accuracy does not necessarily translate to an increase in distribution accuracy. The charts in figures 6.2.29 and 6.2.30 show the results of precise interpolation and 400 point regression respectively.

6.2.69 The precise interpolation shows a deterioration in fit at the lower end of the distribution when compared to the 100 point regression despite having significantly better scenario accuracy. At the upper end of the distribution, the fit is similar to that which results from using 400 point regression.

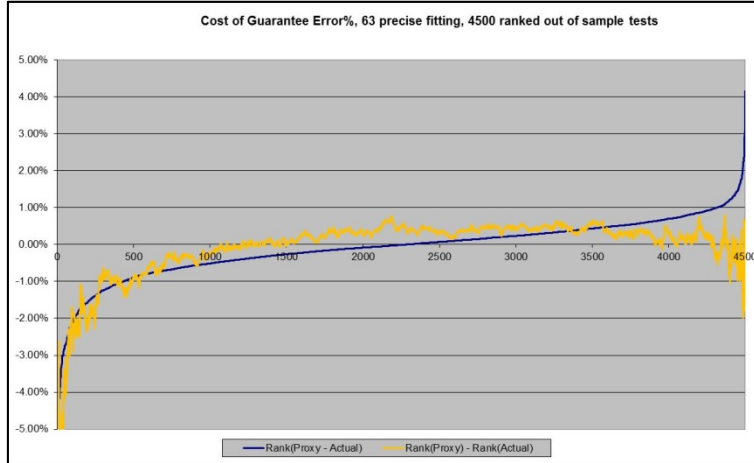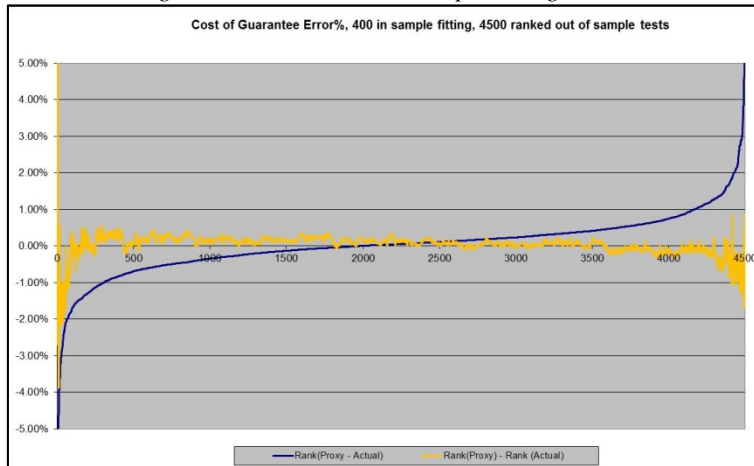*Fig. 6.2.29 – CoG Error, 63 Point Interpolation*


Fig. 6.2.30 – CoG Error, 400 point Regression

6.2.70 For each calibration, the distribution errors become more volatile at the tails of the distribution, oscillating to varying degrees. In fact, the 99.5[th] percentile VaRs, as shown in table 6.2.3, are not the ideal metric for comparing distribution accuracy due to this oscillation. For example, precise interpolation gives the most accurate 1-in-200 VaR but not the most accurate 1-in-100 VaR. The results in table 6.2.4 illustrate.

Table 6.2.4

No. of Calibration Scenarios	100	200	300	400	500	750	1000	63
99.9[th] Percentile Error	1.69%	1.12%	0.77%	0.71%	0.43%	0.36%	0.51%	0.09%
99.5[th] Percentile Error	1.12%	0.98%	0.55%	0.77%	0.75%	0.57%	0.65%	0.29%
99th Percentile Error	0.56%	0.86%	0.69%	0.78%	0.65%	0.74%	0.53%	0.66%
95th Percentile Error	0.57%	0.19%	0.10%	0.26%	0.13%	0.05%	0.01%	0.22%

6.2.71 Finally, we consider the impact on distribution accuracy of increasing the number of out-of-sample test scenarios. This is illustrated by the charts in figure 6.2.31 which show the results of using 4,500 and then 20,000 scenarios for each of the 63 point interpolation and 100 point regression calibrations.

Fig. 6.2.31 – CoG Distribution Accuracy
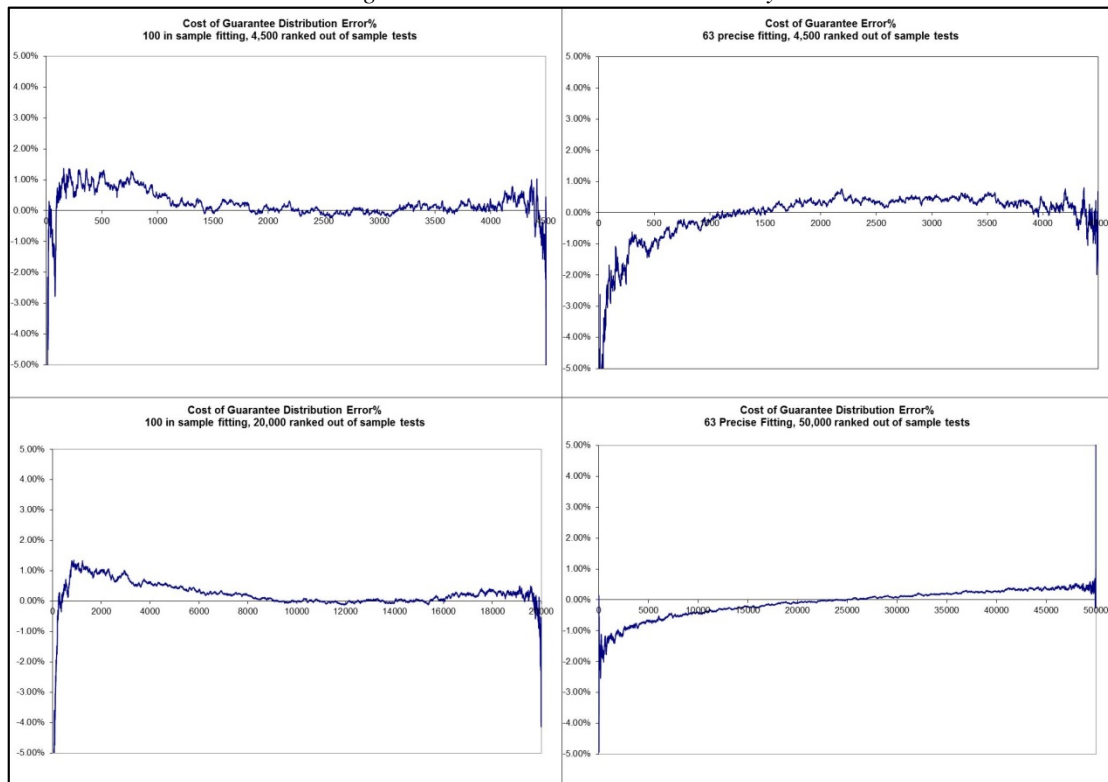
6.2.72 Increasing the number of scenarios to 20,000 acts to smooth out the volatility in the distribution errors whilst retaining the overall shape of the error curve. A further set of tests was performed using 50,000 out-of-sample scenarios which smoothed the curve even further. This observation is consistent with the comments in 4.2 regarding the impact on VaR accuracy of increasing the number of Monte Carlo simulations. In particular, it should be noted from figure 6.2.31 that the 63 point interpolation gives rise to VaR errors at the upper tail of the distribution that range, approximately between +1% and -2% when using 4,500 scenarios whereas using 20,000 scenarios has reduced the impact of simulation error until only the approximation error of around 0.5% remains.

6.2.73 This concludes the case study for replicating polynomials. However, there still remains considerable scope for variation and further analysis, both in the derivation of the polynomial formula and in its calibration and implementation. The purpose of this case study was to give some flavour of the issues that need to be addressed and the decisions that need to be made. It should be obvious that a great deal of subjective judgement was used and hopefully it has been made clear that sometimes these decisions are not in line with, or as would be suggested by, objective application of measured statistics.

6.3 RADIAL BASIS FUNCTIONS

Introduction

6.3.1 Approximation via use of radial basis functions (RBFs) has been used successfully in a number of areas, including image reconstruction in computer graphics. See Holger Wendland (2010) for further background details on RBFs.

6.3.2 An RBF approximates the unknown function f(x) by a function of the form:

$$g(x) = \sum \psi_i \, \Phi(\|x, x_i\|)$$

6.3.3 In this equation, the x_i are the fitting points where the value of the unknown function is known. In the example considered in the later sections, the unknown function is the cost of guarantees and each x_i is the value of the risk drivers (equity, rates, lapse etc.) at that fitting point. The ψ_i are weights assigned to each fitting point. The function $\Phi(\|x, x_i\|)$ is the radial basis function. It is "radial" because its value depends only on the Euclidean distance between the point to be approximated, x, and the fitting point x_i.

6.3.4 The RBFs considered in this section are all interpolations rather than regressions. Hence at each of the fitting points x_i the value of the function to be approximated $f(x_i)$ equals $g(x_i)$. We solve for the weights ψ_i, and so this gives n linear equations in n unknowns. In general, this will have a unique solution and it is possible to choose Φ to ensure that this is the case.

6.3.5 If the function Φ is positive definite, then the system of equations for ψ_i will have a solution. A function   is said to be positive definite if and only if the following two conditions hold:

1. Φ is even, so
2. For all   and all   , for all pairwise distinct   we have

This is a generalisation of the idea of positive definiteness for matrices. See Wendland (2010) for further details and a proof.

6.3.6 The most common choices for Φ satisfy this condition. These choices include:

- **Gaussian.** This has the form $\phi(r) = e^{-(\varepsilon r)^2}$
- **Multi-quadric.** This has the form $\phi(r) = \sqrt{1 + (\varepsilon r)^2}$
- **Inverse Multi-quadric.** This has the form $\phi(r) = 1 / \sqrt{1 + (\varepsilon r)^2}$
- **Thin plate splines.** This has the form $\phi(r) = r^2 \ln(r)$

6.3.7 In general, adding more fitting points will improve the quality of fit. There is a general convergence result that, under mild conditions on the smoothness of the unknown function, ensures that this result holds.

Splines as RBFs

6.3.8 Although the form of the formula in 6.3.2 may be unfamiliar, polynomial RBFs in 1 dimension are a spline. We demonstrate this by way of an example.

6.3.9 Consider the cubic spline illustrated in Figure 6.3.1.

Fig. 6.3.1 – Cubic Spline
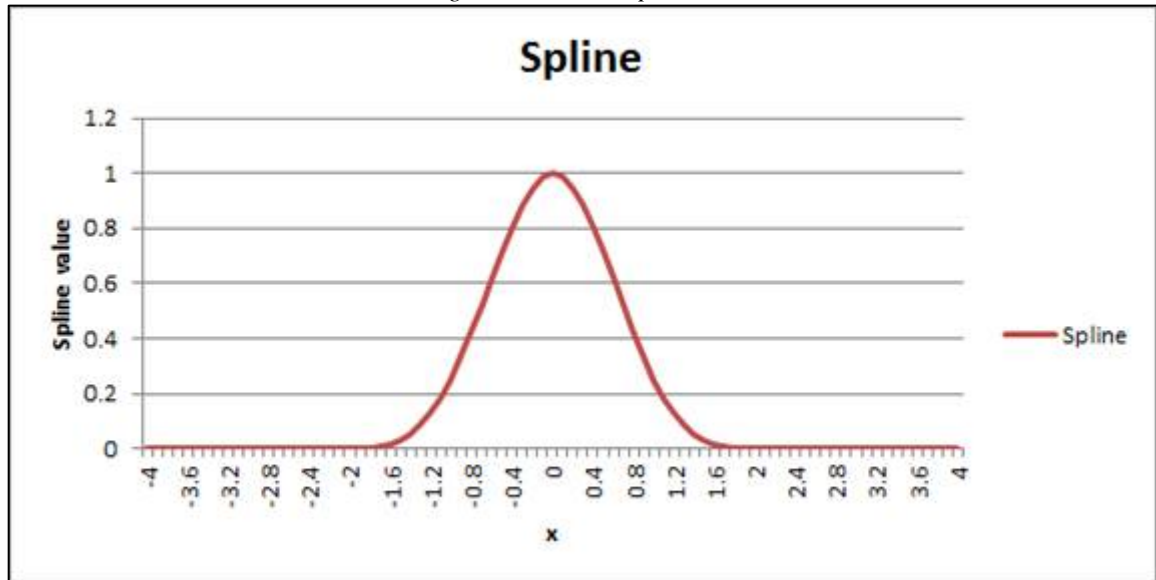
6.3.10 This spline, F(x), is defined as follows:

F(x)	Domain
0	x<-2
¼*(x+2)³	-2≤x<-1
¼*(-3x³ - 6x² +4)	-1≤x<0
¼*(3x³ - 6x² +4)	-0≤x<1
¼*(x-2)³	1≤x<2
0	2≤x

6.3.11 It is possible to re-write F(x) as:

$$F(x) = \sum \quad (\| \quad - \quad \|)$$

where $(r) = r^3$, and the fitting points x_i and associated weights are given by:

x_i	
-2	0.125
0	0.75
1	0.5
2	0.125

This expresses F(x) as an RBF, as required.

Example of RBF Interpolation

6.3.12 In this section, we present the practical results of using RBF interpolation on the with-profits model referred to in [the introduction to this chapter.] The liabilities considered were the cost of guarantees; the asset share; and the total liability, which was the sum of the cost of guarantees and the asset share.

6.3.13 For the majority of this section, the RBF used was multi-quadric, and so the approximation had the form:

$$\sum \psi_i \left(|(|x-x_i|)|^2/\varepsilon^2 + 1 \right)$$

We investigate alternatives to multi-quadric in the final part of this section.

6.3.14 As noted in section 5.3, there are different methods of choosing the fitting points. Throughout this section, we assume the fitting points were chosen from independent normal distributions.

6.3.15 The liability was calculated at 4,400 fitting points in total. The first points were used to calibrate the RBF; and the remaining points were used as out-of-sample test points. The RBF was calibrated with between 50 and 400 fitting points.

6.3.16 The chart in figure 6.3.2 shows the true asset share for the out of sample points plotted against the estimated value when 100 fitting points were used.
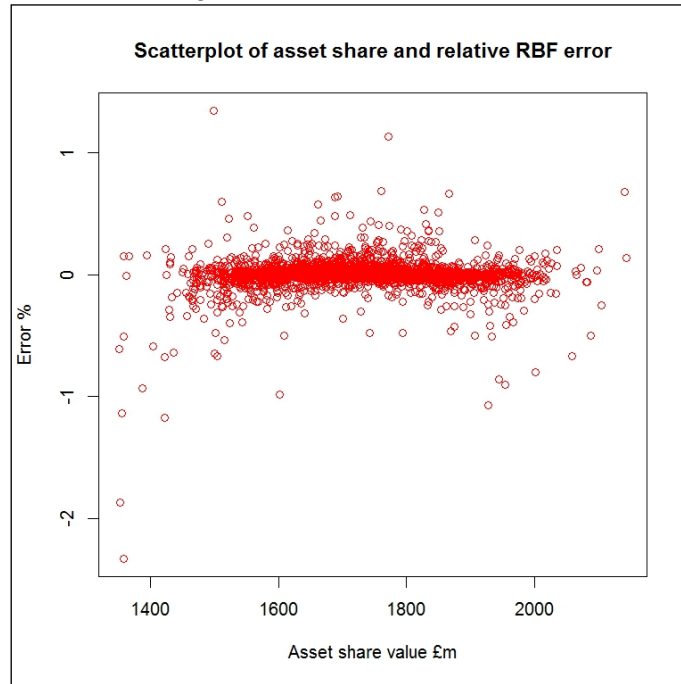
Fig. 6.3.2 – Asset Share, Actual vs. Estimated
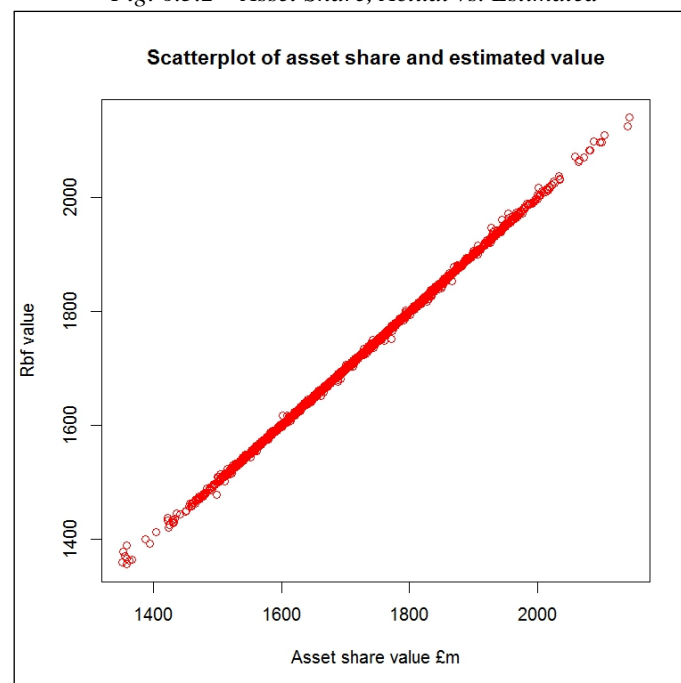
6.3.17 It appears from this diagram that there is a reasonably good fit across the whole domain tested. This is borne out by the chart in figure 6.3.3 which plots the percentage error in the RBF approximation against the true value of the asset share.

6.3.18 It is clear that the maximal absolute percentage error of -2.3% is achieved at the extreme left hand tail of the distribution. In the body of the distribution, the fit is reasonably good. This is borne out by the root mean square error, which equals 0.1%. The 1-in-200 Value-at-Risk estimated by the RBF differed from the true value by 0.3%.

Fig. 6.3.3 – Asset Share, Error %

6.3.19 For the remainder of the section, we consider fitting to the cost of guarantees, which are more non-linear and therefore a more interesting case study than asset share

*Fig. 6.3.4 – Cost of Guarantees, Actual vs Estimated*
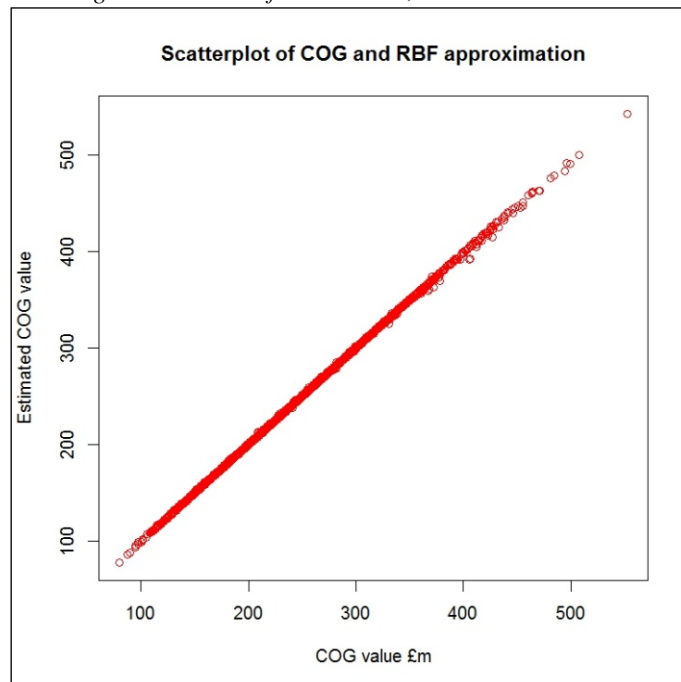
6.3.20 The chart in figure 6.3.4 shows the true cost of guarantees plotted against the estimated values for the out-of-sample test points when there were 100 fitting points. This shows that there was a reasonably good fit from the approximation, although visually it appears to be a worse fit than asset share.

6.3.21 The chart in figure 6.3.5 shows the percentage error of the fit, plotted against the true value of the COG. The chart indicates that although there is generally a good fit for most value of the COG, there are some outliers and the fit is generally poorer towards the extreme ends of the distribution.

6.3.22 The mean square error of the fit is 0.3%. The 1-in-200 Value-at-Risk estimated by the RBF differed from the true value by 0.4%. The maximum error was 3.5% and the minimum error was -2.1%.

Fig. 6.3.5 – COG Error %, 100 fitting points
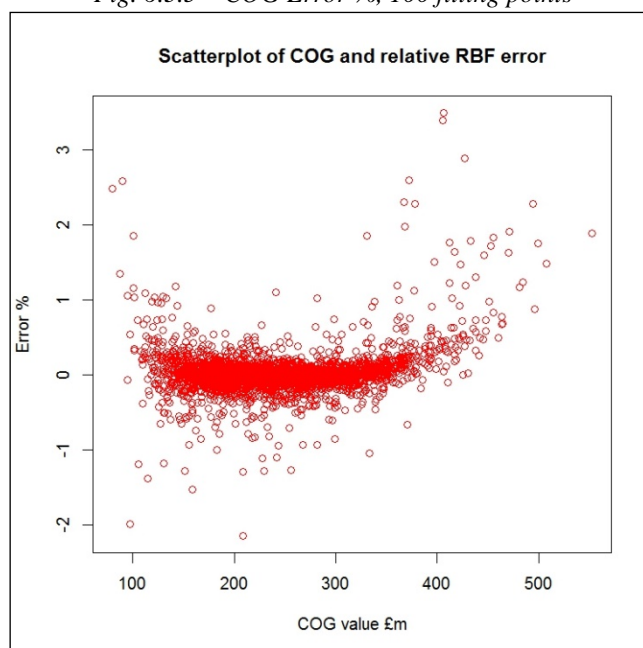
6.3.23 We now investigate the sensitivity to different numbers of fitting points. Under the assumption that the liability is reasonably well behaved, extra fitting points should improve the quality of fit. This is borne out in this example. Table 6.3.1 shows the statistics of using 400 fitting points compared to using 100.

Table 6.3.1

Statistic	100 Fitting Points	400 Fitting Points
Root means square error	0.3%	0.1%
Error in 1-in-200 VaR	0.4%	0.0%
Maximum error	3.5%	1.8%
Minimum Error	-2.1%	-0.9%

6.3.24 The chart in figure 6.3.6 shows the chart of the true COG plotted against the percentage error using RBF interpolation with 400 fitting points. Visually, the

fit appears to be an improvement on the result with 100 fitting points. It is worth noting that the error in estimating the 1-in-200 VaR is in line with the root mean square error and not in line with the maximum, minimum or 1-in-200 level of the error. This is in line with the result discussed in section 4.2 and demonstrated in 6.2 (replicating polynomials).

Fig. 6.3.6 – COG Error %, 400 fitting points
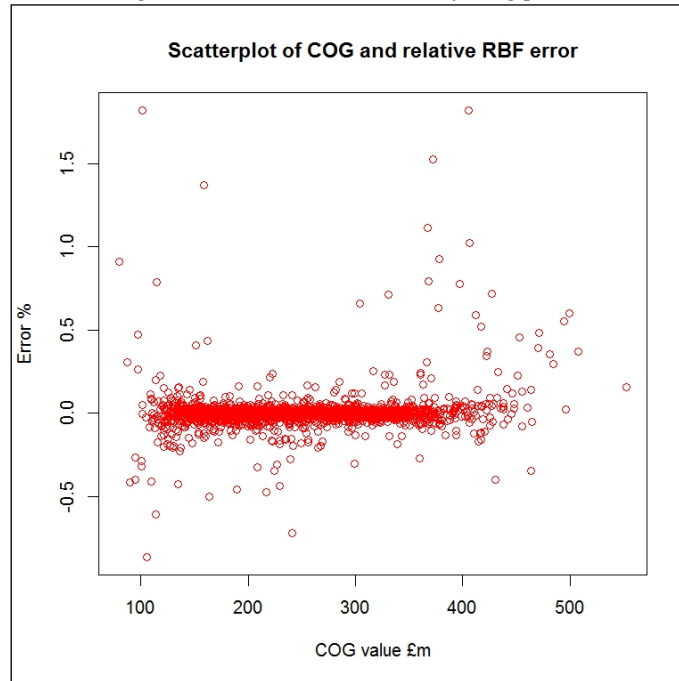
6.3.25 A formal investigation of goodness of fit would use some criterion such AIC. See Section 4.2 for a description of AIC.

Fig. 6.3.7 – COG Error &, 100 fitting points
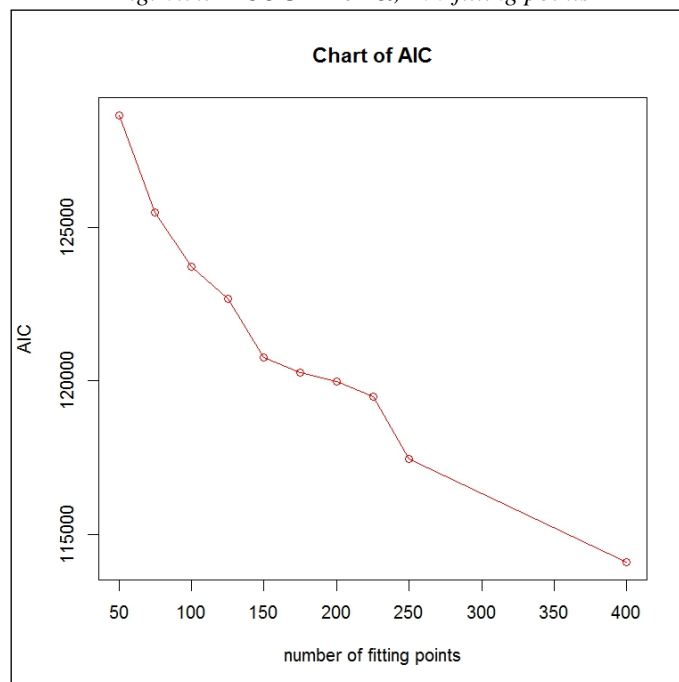
6.3.26   The chart in figure 6.3.7 shows the AIC of the fit achieved with various numbers of fitting points. We investigated the RBF model with 50; 75; 100; 125; 150; 175; 200; 225; 250 and 400 fitting points. A lower AIC indicates a better fit. As expected, increasing the number of fitting points improves the fit as measure by AIC.

6.3.27   This chart demonstrates the trade-off between complexity and accuracy. More fitting points result in a better fit to the unknown variable. However, offsetting this, if a greater the number of fitting points is chosen, the model is more complex with respect to calibrating, running and validating the heavy model to produce the fitting point and out-of-sample results.

6.4   REPLICATING PORTFOLIOS

Introduction

6.4.1   A replicating portfolio is a pool of assets designed to approximately reproduce the market values of a pool of liabilities across a large number of stochastic scenarios. These can then be used to create an ideal hedge for the liabilities, which a firm then has the option to purchase in the capital markets. Also, an important by-product of a replicating portfolio is to act as a proxy for the liabilities, which (given certain conditions hold) can be used to predict the behaviour of change in value of the liabilities based on a number of economic outcomes.

6.4.2   In this paper, we shall focus the majority of the discussion on the second use (as a proxy model). As with other proxy models, the single biggest advantage compared to 'heavy' models is simply, speed. This is made possible by a number of 'closed form' formulae available from advances in financial mathematics (e.g. Black Scholes formula for equity options and Black formulae for swaptions). Even where closed form formulae do not exist, the nature of the replicating assets is such that they can be recalculated significantly quicker than the heavy models they are intending to replicate.

6.4.3   The range of assets used depends on the aim of the replicating portfolio. If required to calculate financial hedges (or potential hedges), the family of assets are often restricted to vanilla instruments whose prices can be found in the capital markets. However, in theory, there is no requirement to restrict the family of replicating assets to those that can actually be purchased, and a number of 'exotic' options / mathematical constructions can be used. This is where a replicating portfolio merges with some of the other proxy models described in this paper and so won't be discussed further in this chapter.

6.4.4   There are two main advantages to a replicating portfolio:

Intuition
6.4.5   The replicating portfolio method scores very highly on the intuition front, especially if the senior management of a firm have significant investment knowledge. For example, there are clear analogies between a with profits

product with a given term and guaranteed amount to a put option with a given strike and maturity date. Furthermore, the strike of the replicating option provides valuable information to the insurer on the 'moneyness' of the guarantees that they have provided to the customers.

Use for Hedging

6.4.6   Although a fairly obvious use, it is worth bearing in mind. In particular, given a theoretical replicating portfolio, the firm often has a number of choices in terms of what risks it wants to hedge and what risks it wants to retain based on cost benefit analysis. The replicating assets could be decomposed into 'greeks' (the rate of change of replicating portfolio values to key market variables), and a firm could decide to selectively hedge some of the 'greeks'. A classic real life example is using replicating portfolios to dynamically delta-hedge the liabilities (i.e. hedge the first order sensitivities to equity market movements, and rebalance frequently). Note that the use of greeks such as delta, gamma, vega and rho are coming into more common usage in terms of insurance risk monitoring systems, in colloquial use if not for actual hedging activity!

Determining Formula Structure

6.4.7   An important design choice within the replicating portfolio method is the choice of replicating assets to use. Unfortunately, unlike replicating polynomials, there is no guaranteed 'one size fits all' method, and considerable thought needs to be given to the family of assets used. This does very much depend on the nature of the design of the heavy model, as well as details about the product / liability under consideration. For example, with profits products would have a number of distinctive features such as term, nature of guarantees, level of guarantees and cohort, which may provide insight into the opportunity set of replicating assets to use. Likewise, products that behave more like annuity would have fundamentally different cashflows and features.

6.4.8   Once the 'formula structure' has been designed (i.e. the family of assets to use as the opportunity set), the coefficients boil down a rather large regression and optimisation exercise. This requires similar numerical and computing challenges as some of the other basis functions described in this paper. In addition, there may need to be a number of iterations of replicating portfolios, progressively moving from the exotic/theoretical families to the more practical vanilla instruments that a firm may want to trade in the capital markets.

Limitations of Replicating Portfolios

6.4.9   Replicating portfolios can help insurance companies improve the speed and accuracy of certain calculations. However, replicating portfolios have some limitations.

Non-financial risks

6.4.10 Replicating portfolios as we know them can be very useful for quantifying and managing financial risk, but not ideal for non-financial risks. It is possible to merge a replicating portfolio with some of the other basis functions in proxy modelling to mitigate this aspect. However, in practice it means that when there is a material change in liability methodology and assumptions (e.g.

decrements, management actions, etc...) there would be a need to recalibrate the replicating portfolio.

Know-How
6.4.11 Replication work requires thorough knowledge of both sides of the balance sheet, as well as the interaction of assets and liabilities. This is not usual, as the common path is for the Actuarial department of an insurance firm to specialise in the understanding of the liabilities, and for the investment management side of an insurance firm to specialise in knowledge of the assets! Additionally, the introduction of stochastic calculus is (relatively) new in the actuarial syllabus, having been introduced circa 2004, and replicating portfolio theory remains at present remain a specialist area of knowledge outside of the Actuarial (or CFA) syllabus. Thus a firm has the difficult choice of either building a critical mass of individuals with quantitative skills as well as knowledge of both sides of the balance sheet to build a tool, or to outsource (at a considerable cost!) to consultants.

Knowledge of the Liabilities
6.4.12 Finally, given the sheer number of regression possibilities when creating the family of assets representing replicating portfolios, these require detailed knowledge of the firm's products in order to succeed.

6.5 COMMUTATION FUNCTIONS

6.5.1 The Commutation Proxy model is not prevalently used in industry. It is presented here as a methodology covering the advantages of replicating portfolio methodology with respect to marketable securities but expanding that methodology further to include non-market related risk drivers. Thus any basis function (as describe in Section 4) used in a replicating portfolio proxy model is also available in a commutation proxy model.

What is a commutation function?

6.5.2 Commutation functions were developed before computers were widely used in insurance. Without going into the exact definitions, essentially they are a series of vectors used to describe aspects of an insurance product, where the index of the vector is over time.

6.5.3 With classic commutation functions, individual risk are projected separately and combined together in a dot product formula. For example, for life insurance, you might have:

- a Nominal Amount life cover vector $\vec{N} = (n_1, \ldots, n_{Last})$ with indices indicating years forward,
- a Discount Factors vector $\vec{D} = (D_1, \ldots, D_{Last})$,
- a Survival vector $\vec{l} = (l_1, \ldots, l_{Last})$, and
- a Persistency vector $\vec{P} = (per_1, \ldots, per_{Last})$.

6.5.4 The present value of this situation is given by the equation:

$$PV = \sum_{i=1}^{Last} n_i \cdot d_i \cdot l_{\cdot i} \, per_i \, .$$

6.5.5 A generalized Commutation (our usage, not common usage) may use something other than a dot product. It is a function of vectors:

$$PV = f(\vec{N}, \vec{D}, \vec{l}, \vec{P}) \text{ for some function } f.$$

6.5.6 Clearly, the classic commutation function is a special case of the general one. It is possible to have vectors of different size in the function. For example:

$$\text{Discounted Cash flows: } PV_{Commutator} = \sum_{i=1}^{Last} \frac{CF_i}{(1+r)^i} = f(\vec{CF}, r)$$

6.5.8 This is a standard actuarial formula, which is a generalized commutation. The cash flows naturally form a vector, $\vec{CF} = (c_1, \ldots, c_{Last})$. The variable r is viewed as a vector of length 1.

6.5.9 The power of this methodology rests on how well these generalised commutation functions can be defined. Clearly the discount function, $\vec{D} = (D_1, \ldots, D_{Last})$, can be defined as a function of interest rates. The vector might be 30 years long, but it may be a function of only 3 interest rates variables, as would be the case in a model that used Principal Component Analysis to develop a 3 factor model. If a spread-laden rate were needed, again the final vector could be as long as our projection period but the number of variables would be much smaller.

6.5.10 The general idea is that the above expression for *PV* would be our new basis function as presented in Section 3.1. It would be a function of the underlying risk drivers, whatever they are.

6.5.11 The commutation functions would have a strong advantage over polynomials, however, because the form of the basis function is actuarially/financially based, and accordingly should better reflect the shape of the actual function of risk drivers. Expressed mathematically, the commutation function would extrapolate and interpolate better away from the scenarios points from the heavy lifts better than polynomial functions would.

6.5.12 Commutation functions could include all the function from replicating portfolio proxy models, but also include functions of risk drivers other than market risk drivers, such as pure insurance drivers (mortality, expenses, lapses, etc.) and as such would reflect these risk drivers.

7   SUMMARY

7.1   FINAL THOUGHTS

7.1.1   The interest in, and use of, proxy models within the insurance sector has grown considerably in recent years driven by both regulatory (Solvency II) and risk management requirements. Arguably, some organisations did not have as much time as they would have liked to review the proxy modelling landscape before choosing a solution to be embedded in their risk management framework. The imperative for these organisations is likely to be the understanding and improvement of their models. However, in order to achieve this it may be necessary to take a step or three backwards and recognise the fundamental choices that have already been made in selecting a particular model and with that choice, the potential limitations. Understanding those limitations will be crucial to devising ways over or around them. For those organisations still contemplating proxy models, the situation is perhaps a little better as there is opportunity to learn from the experience of others. However, the required knowledge and information is the same.

7.1.2   If the question being asked now is "Which is the best proxy model?" then unfortunately there is no right answer. Proxy modelling, by its nature, requires a series of compromises to be made, in both the abilities and features of the model; some features will be preferred to others, accuracy in one metric maintained at the expense of others. As such, the chosen model (and design) should be driven by the uses to which the model will be put. Perhaps the question should be expanded to ask "Which is the best proxy model to answer this particular set of questions?" whatever those question may be. If we consider the natural desire and tendency to expand the uses of, and therefore extract maximum value from, a model then perhaps the question should instead relate to maximum flexibility or adaptability.

7.1.3   Unfortunately, these questions are no more easily answered and, this being a working paper, we have not attempted to offer a solution here, merely a discussion around the issues to be considered and perhaps a basic framework for comparison. Likewise, detailed conclusions are deferred until later publication.

7.2   NEXT STEPS

7.2.1   Undoubtedly, as knowledge and experience of proxy models increase then so the techniques being used will be refined and enhanced to take advantage of new information. Technology, too, will play a part leading to increases in efficiency and speed which will offer the opportunity for more calibration scenarios and increasingly complex formula structures. We can easily imagine an increasing reliance on automated processes for determining formula structure. Strides have already been made in this direction with techniques such as those employed by Least Squares Monte Carlo. A continuation of this development through a greater understanding and increased use of intelligent decision making algorithms will likely prove necessary.

7.2.2   At the same time, judgement shall remain an important aspect of the process, but perhaps the pure judgement would be concentrated around the big-picture model design aspects. This will be especially true as more insurers have the luxury, time and resource when being tasked with building a proxy model from a clean sheet of paper (as opposed to being forced to choose a set menu template to comply with 'impending' capital regulations).

7.2.3   If we consider the development of liability models from analytic functions and formulae through to stochastic cashflow models, in a strange way we have now come full-circle with a return to functions and formulae. The irony, however, is that the formulae being used now are often less sophisticated than those originally replaced by cashflow models two decades ago.

7.2.2   Perhaps then, there are lessons to be learned from the past. Recognising the disconnect between that which we are trying to model and the basis functions being used to do so, there seems considerable scope for development of more bespoke basis functions, many of the principles and methods discussed in this paper being equally applicable to more bespoke functions, particularly those that better reflect the characteristics of the heavy model under consideration. Marrying the old with the new, therefore, would certainly appear to be an area for further discussion and research.

REFERENCES

Frankland et al. (2013). "Difficult Risks and Capital Models – A Discussion Paper", *Presented to The Institute and Faculty of Actuaries 22nd April 2013.*

P Glasserman (2003). *"Monte Carlo Methods in Financial Engineering (Stochastic Modelling and Applied Probability"*, Springer.

C. Hursey & R. Scott (2012). "Replicating Formulae – Efficient Calibration Techniques", *Presented to the Staple Inn Actuarial Society, 7th February 2012.*

Li, Y.M. (n.d.). "Least Squares Approximation by Legendre Series", www.infogoaround.org/JBook/LSQ_Legendre.html [accessed Feb 2014]

F. A. Longstaff and E. S. Schwartz (2001). "Valuing American Options by Simulation: A Simple Least-Squares Approach", *The Review of Financial Studies, Spring 2001* **Vol. I4. No. I**, pp. 113-147.

L Stentoft (2004) "Convergence of the Least Squares Monte-Carlo Approach to American Option Valuation", *Management Science,* **50 (9)**, pp. 1193-1203

M. H. Stone (1948). "The Generalized Weierstrass Approximation Theorem", *Mathematics Magazine* **21 (4):** pp 167–184

H. Wendland (2010). "Scattered Data Approximation", *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press.

APPENDIX 1 – CONVERGENCE OF LEAST SQUARES MONTE CARLO

APPENDIX 2 – OPTIMISED COMPONENTS VS OPTIMISED WHOLE, AN EXAMPLE

A.2.1 We start by considering some unknown function in two variables. We assume that this function can be approximated with arbitrarily small error using a two-factor 4th order polynomial $Q(x,y)$, i.e. 5th order terms are negligible.

$$
\begin{aligned}
Q(x,y) = \ &a_{24}x^4y^4 + a_{23}x^4y^3 + a_{22}x^4y^2 + a_{21}x^4y + a_{20}x^3y^4 \\
&+ a_{19}x^2y^4 + a_{18}xy^4 + a_{17}x^3y^3 + a_{16}x^3y^2 + a_{15}x^3y \\
&+ a_{14}x^2y^3 + a_{13}x^2y^2 + a_{12}x^2y + a_{11}xy^3 + a_{10}xy^2 \\
&+ a_9xy + a_8x^4 + a_7x^3 + a_6x^2 + a_5x + a_4y^4 + a_3y^3 \\
&+ a_2y^2 + a_1y + a_0
\end{aligned}
$$

Optimised Whole

A.2.2 Assuming we wish to approximate using a two-factor 3rd order polynomial $p(x,y)$, the optimal fitting points are given by the intersections of the roots of the 4th order Legendre Polynomials. This will give 16 fitting points such that the interpolating polynomial will be the best fit surface in the least squares sense.

A.2.3 The diagram in figure A.2.1 shows a bird's eye view of the error surface that results from finding the least squares fit across a large number of data points. The intersection between the two differently shaded areas indicates zero error. It can be shown that zero error always occurs at the intersection of the roots of the Legendre polynomials as shown by the thick black lines. This indicates that they provide optimal fitting points.

Fig. A.2.1
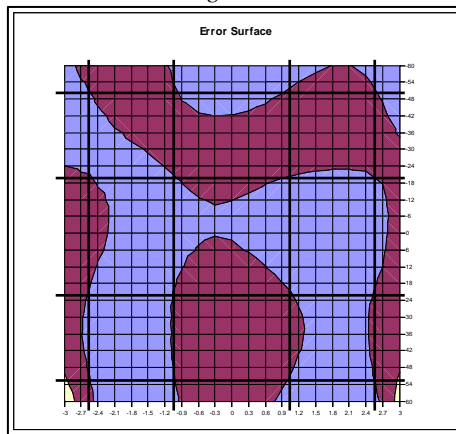
A.2.4 The replicating formula that results is a two-factor 3rd order polynomial which can be deconstructed into its component parts:

$$p(x,y) = f(x) + g(y) + h(x,y)$$

Where *f* and *g* are cubic functions in one variable representing the marginal risk functions and *h* represents the non-linearity function consisting of cross terms only, e.g. xy, x^2y, xy^3. Slight adjustment is need to the combined

formula due to the effective doubling counting of the constant terms from each of the marginal risk functions:

$$p(x,y) = f(x) + g(y) + h(x,y) - a_0$$

$$f(x) = b_7x^3 + b_6x^2 + b_5x + {}_xb_0$$

$$g(x) = b_3y^3 + b_2y^2 + b_1y + {}_yb_0$$

$$\begin{aligned} h(x,y) = {} & b_{17}x^3y^3 + b_{16}x^3y^2 + b_{15}x^3y \\ & + b_{14}x^2y^3 + b_{13}x^2y^2 + b_{12}x^2y \\ & + b_{11}xy^3 + b_{10}xy^2 + b_9xy \end{aligned}$$

$$b_0 = {}_yb_0 + {}_xb_0 - a_0$$

$$\begin{aligned} p(x,y) = {} & b_{17}x^3y^3 + b_{16}x^3y^2 + b_{15}x^3y + b_{14}x^2y^3 + b_{13}x^2y^2 \\ & + b_{12}x^2y + b_{11}xy^3 + b_{10}xy^2 + b_9xy + b_7x^3 + b_6x^2 \\ & + b_5x + b_3y^3 + b_2y^2 + b_1y + b_0 \end{aligned}$$

A.2.5 If we plot *f* and *g* we get the two charts in figures A.1.2 respectively. It can be seen that the resulting marginal risk functions are not optimal so care must be taken when using the components of a holistically derived approximation. For example, using the marginal risk functions to analyse sensitivity to individual risks may lead to incorrect conclusions, especially at the tails.
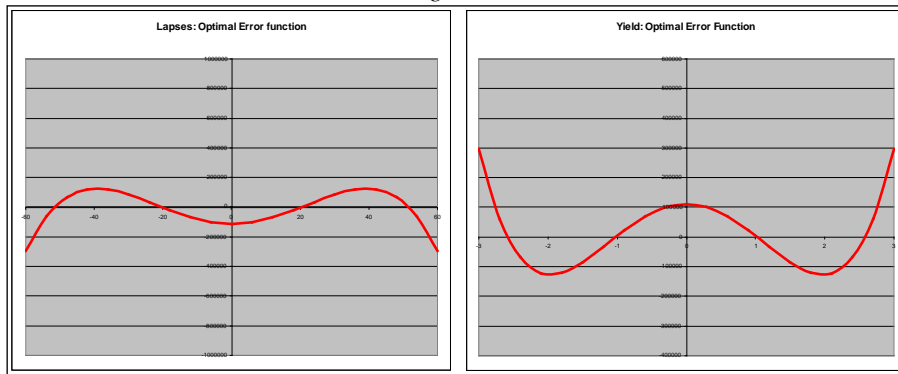
Fig. A.2.2
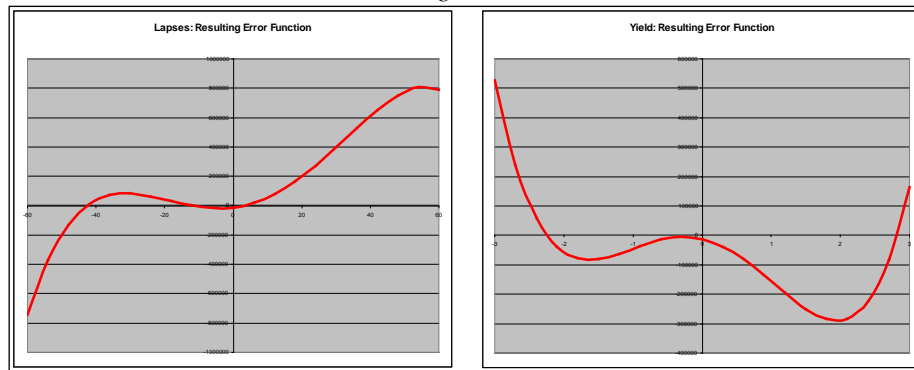
Optimised Components

A.2.6 We now consider the same example but start by finding the optimal cubic approximations for the marginal risks, f(x) and g(y), first. The two diagrams in figure A.2.3 show the resulting error curves. These curves have zeroes at the roots of the 4th order Legendre polynomial.

A.2.7 Errors are reduced significantly and more importantly have a predictable shape given their derivation from Legendre polynomials.

A.2.8 The optimal non-linearity approximation for h(x, y) is found by fitting to the intersection of the one dimensional solutions where the error is constrained to be zero along the axes, i.e. not at the roots of the Legendre polynomials.

Fig. A.2.3

A.2.9   Combining all three optimal components we obtain the final replicating formula and can compare its error surface with that derived from the holistic optimal fit as shown in the two diagrams in figure A.2.4.

Fig A.2.4
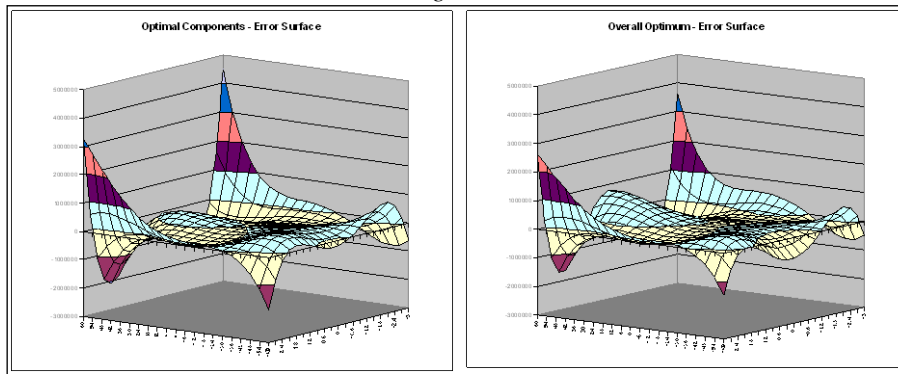
A.2.10 There is some loss of goodness of overall fit when using optimal components, though this seems to be limited to the extreme corners of the domain. Depending on what use the formula is to be put, this loss in quality of overall fit may or may not be desirable.

Further Analysis

A.2.11 We can analyse the functions further to gain an understanding of the factors that may influence to what extent a holistic approach may lead to unsuitable marginal risk approximations.

A.2.12 We can derive functions for each of the approximation function coefficients in terms of the coefficients of the function we wish to approximate and the domain over which the best fit is required. Table A.2.1 provides these functions in this example, where a 3rd order fit is required for a 4th order function. We then summarise these results in terms of the dependencies in Table A.2.2.

Table A.2.1

		Optimal Components	Holistic (overall optimal)
Constant	b_0	$a_0 - 3/35a_8R_x^4 - 3/35a_4R_y^4$	$a_0 - 3/35a_8R_x^4 - 3/35a_4R_y^4 + (3/35)^2a_{24}R_x^2R_y^2$
Coefficients of y	b_1	a_1	$a_1 - 3/35a_{22}R_x^4$
	b_2	$a_2 + 6/7a_4R_y^2$	$a_2 + 6/7a_4R_y^2 - 3/35a_{21}R_x^4 - (3/35)(6/7)a_{24}R_x^4R_y^2$
	b_3	a_3	$a_3 - 3/35a_{23}R_x^4$
Coefficients of x	b_5	a_5	$a_5 - 3/35a_{18}R_y^4$
	b_6	$a_6 + 6/7a_8R_x^2$	$a_6 + 6/7a_8R_x^2 - 3/35a_{19}R_y^4 - (3/35)(6/7)a_{24}R_x^2R_y^4$
	b_7	a_7	$a_7 - 3/35a_{20}R_y^4$
Coefficients of non-linearity	b_9	a_9	a_9
	b_{10}	$a_{10} + 5/7a_{18}R_y^2$	$a_{10} + 6/7a_{18}R_y^2$
	b_{11}	a_{11}	a_{11}
	b_{12}	$a_{12} + 5/7a_{21}R_x^2$	$a_{12} + 6/7a_{21}R_x^2$
	b_{13}	$a_{13} + 5/7a_{19}R_y^2 + 5/7a_{22}R_y^2 + (5/7)^2a_{24}R_x^2R_y^2$	$a_{13} + 6/7a_{19}R_y^2 + 6/7a_{22}R_y^2 + (6/7)^2a_{24}R_x^2R_y^2$
	b_{14}	$a_{14} + 5/7a_{23}R_x^2$	$a_{14} + 6/7a_{23}R_x^2$
	b_{15}	a_{15}	a_{15}
	b_{16}	$a_{16} + 5/7a_{20}R_y^2$	$a_{16} + 6/7a_{20}R_y^2$
	b_{17}	a_{17}	a_{17}

Table A.2.2

	Optimal Components	Holistic (overall optimal)
Coefficients of marginal approximation function in x	Function of the coefficients of the underlying marginal function in x and the domain in x	Function of the coefficients of the underlying marginal function in x, the coefficients of the underlying non-linearity function and the domains in both x and y
Coefficients of marginal approximation function in x	Function of the coefficients of the underlying marginal function in y and the domain in y	Function of the coefficients of the underlying marginal function in y, the coefficients of the underlying non-linearity function and the domains in both x and y
Coefficients of non-linearity approximation function in x and y	Function of the coefficients of the underlying non-linearity and the domains in both x and y	Function of the coefficients of the underlying non-linearity and the domains in both x and y

Summary

A.2.13 When the optimal fit is found holistically, coefficient of the marginal approximation functions in x and y are functions of the coefficients of non-linearity in the original function and functions of the domains in both x and y:

- Adjusting the range in y will change the marginal fit to the function in x as will a change in non-linearity.

- The more severe the non-linearity, the further from optimal the marginal approximation functions become

A.2.14 When optimal components are found, the coefficients of the approximation functions in x and y are functions of the coefficients of the corresponding underlying marginal risk functions and corresponding domain only:

- Adjusting the range in y will have no impact on the marginal fit to the function in x nor will the extent of underlying non-linearity
- The more severe the non-linearity, the further from optimal the overall fit becomes.

A.2.15 As a final observation, it should be noted that this example has used a 3rd order approximation function in order that the optimal component fit is different from the fit derived through optimising the whole. For even-ordered approximation functions, the optimised component fit is the same as that derived through optimising the whole. This is due to the optimal non-linearity surface being zero along the x and y axes. Only odd-powered Legendre polynomials have roots at zero meaning that when optimising the whole only even-powered approximation functions will provide a non-linearity component that is zero on the x and y axes. For odd ordered approximations, the non-linearity surface will be non-zero along the x and y axes.

APPENDIX 3 – WEIGHTED LEAST SQUARES FIT, SAMPLE PROOF

Institute
and Faculty
of Actuaries