The Actuarial Profession
making financial sense of the future

Life Conference and Exhibition 2011
Stuart Carroll, Christopher Hursey

Efficient Curve Fitting Techniques

20-22 November

Agenda

- Background
- Outline of the problem and issues to consider
- The solution
- Theoretical justification
- Further considerations
- Practical issues
- Outcome
- Questions or comments

1

Background

- We wished to use an internal model running in excess of one million scenarios
  - Many different products
  - And many different risk factors
- It was not feasible to evaluate liabilities accurately in every scenario
- Therefore the decision was taken to formula fit liability values
  - Use polynomial approximation formulae
  - Fitted to a limited number of evaluation points
- We need to capture all material risk dependencies
  - Each risk factor leads to a marginal risk function in one variable
  - Non-linearity between risk factors leads to non linearity functions in two or more variables
  - Marginal risk functions and non linearity functions are summed to give the final approximation formula.

2

Outline of the problem and issues to consider
Requirements

- Repeatable process that is objective
- Measurable performance
- Risk management exercise, not just compliance
  - We wish to model the full distribution
- Sufficient number of sample points for accuracy
  - Implies more sample points
- Reasonable run times
  - Implies fewer sample points
- Need to resolve the tension between the previous two points.

3

Outline of the problem and issues to consider
Testing and Validation

- Measures of goodness of fit
  - Least squares
  - Maximum absolute error
  - Error dependency
  - Error Bias
- Out of sample testing
  - How many to be statistically significant
- Analysis of change
- Understanding the results.

4

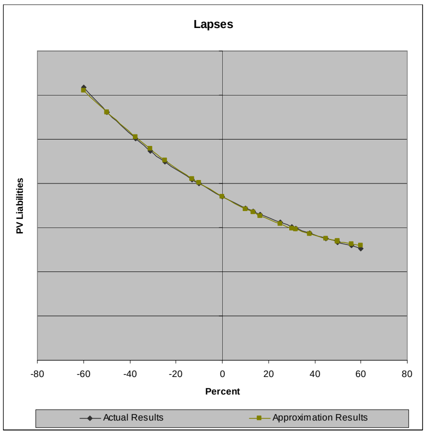
Outline of the problem and issues to consider
Other considerations

- Determining the limits of the model
- Range for each risk factor
  - Allow for movements in roll-forward
  - Four standard deviations?
- What about plan B?
  - Must distinguish between analysis phase and production phase
  - What if goodness of fit fails once in production?

5

The Solution
A simple example

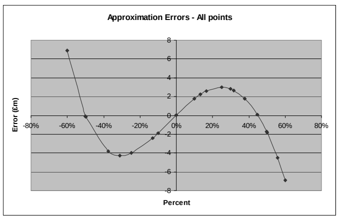
- Consider the variation in liability value with respect to a single risk factor
- The least squares quadratic is fitted using all evaluation points
- Initial fit is reasonable but uses too many calculations to be feasible for the production phase
- How do we reduce the number of fitting points?

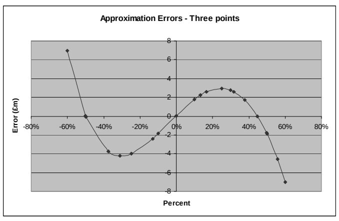


6

The Solution
The Error Curve

- The error curve is cubic
- Three roots at three points of intersection
- Those three points define a unique quadratic function
- Fit to the three points of intersection only
  - same approximation function
  - same error curve
- No loss in performance for greatly reduced effort.





7

The Solution
Implications

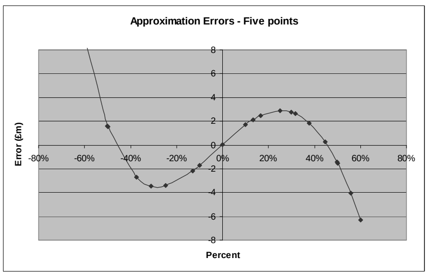
- In general, we wish to approximate our unknown function by an (n-1)-order polynomial
- The least squares (n-1)-order polynomial approximation will intersect the unknown function n times
- If those n points of intersection can be determined, they provide the optimal fitting points, or “nodes” for our approximation function
- We can achieve as good a fit using n nodes as using an infinite number of nodes
- Similarly, we can determine the points of maximum error
  - Allows us to estimate maximum error
  - more powerful than estimating maximum sample error.

8

The Solution
More points is not necessarily better

- What happens when we add points to try and improve the fit?
- The error curve changes
  - Maximum error has increased
  - Sum squared error is worse
- Adding more points has led to a deterioration in the fit

Why?



9

The Solution
Rationale

- This is not a regression problem but an approximation problem
- For a given order of polynomial approximation function there exists a unique optimal error curve, i.e. one that deviates least from zero in the least squares sense
- Identifying solutions to the optimum error curve allows us to use the absolute minimum number of nodes that are required to uniquely define the optimum approximation function
- Adding further nodes shifts the error curve and, by definition, results in a sub-optimal approximation.

10

The Solution
Determining the fitting points

- We introduce a single assumption: our unknown function can be accurately represented by an n-order polynomial
- Powers of (n+1) and above are vanishingly small
- Under this assumption the n points of intersection depend only on the range, or “domain”, over which the least squares best fit is performed
- Similarly, the turning points, or points of maximum error, also depend only on the domain
- Usefully, all our points of interest are now independent of the unknown function and can be determined analytically.

11

The Solution
An example

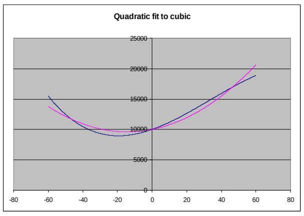
- We wish to approximate a liability function with a quadratic (n=3)
- Assuming powers of four and above are immaterial, we are attempting to approximate an unknown cubic
- the least squares quadratic best fit to any cubic, intersects that cubic at three points

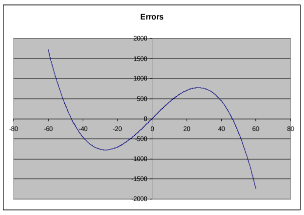
$$P_1 = \frac{(L_2 + L_1)}{2} \qquad P_{2,3} = \frac{(L_2 + L_1)}{2} \pm \sqrt{\frac{3}{5}}\frac{(L_2 - L_1)}{2}$$

and has turning points

$$M_{1,2} = \frac{(L_2 + L_1)}{2} \pm \frac{1}{\sqrt{5}}\frac{(L_2 - L_1)}{2}$$

where L_1 and L_1 represent the upper and lower limits of the domain.

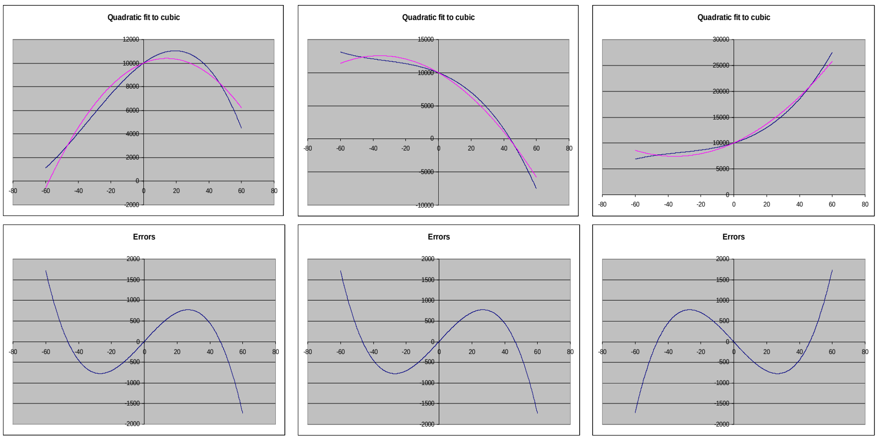




12

The Solution
More examples

Quadratic least squares fit using 1,200 points and the resulting error curves



Points of intersection do not depend on the function.

13

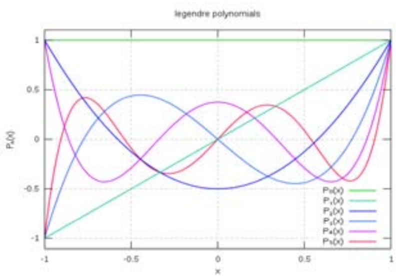
The Solution
Implications

- We can make an assumption about the order of polynomial that accurately represents the unknown function
- Given this assumption, the nodes that give optimum least squares fit can be determined without any prior analysis or knowledge of the unknown function
- Goodness of fit tests will fail for one of two reasons
  - a good enough fit is not possible, or
  - the initial assumption is invalid
- In either case. simply revise the assumption and try again
- As long as the initial assumption holds, once identified, the fitting nodes and points of maximum error remain fixed.

14

Theoretical Justification

- Weierstrauss Approximation Theorem - Any continuous function on a closed and bounded interval can be approximated on that interval by polynomials to any degree of accuracy
- The optimum nodes can be shown to correspond to the roots of the Legendre polynomials for all n



- Properties of Legendre polynomials
  - Orthogonal
  - Series convergence on [a, b]
  - Minimum deviation from zero
- We effectively use a truncated Legendre series

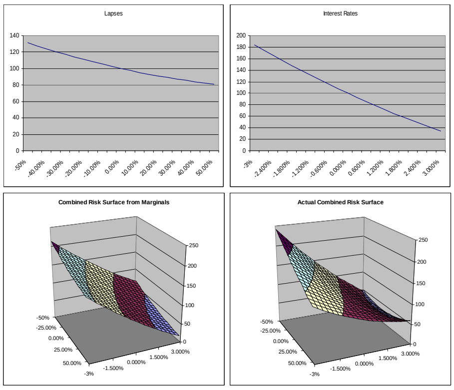
$$f(x) - \sum_{k=0}^{n-1} a_k L_k(x) = a_n L_n(x) + \varepsilon$$

- Forces the error curve to be the n-order Legendre polynomial.

15

Further Considerations
Non Linearity

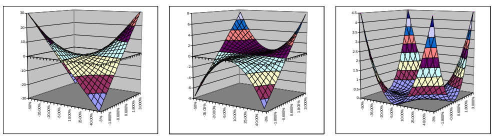
- The same theory can be extended and applied to non linearity functions in two or more variables
- Construct a combined risk surface by adding marginal risk functions
- Compare with actual combined risk surface to evaluate non linearity
- If there is no non linearity between two risk factors then there should be no cross terms in the approximation formula.



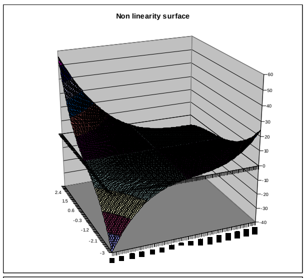
16

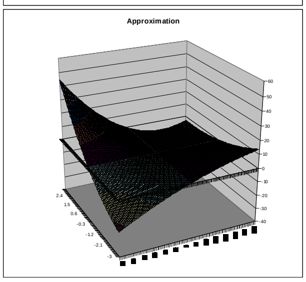
Further Considerations
Non Linearity

- Non-linearity is the difference between the combined impact of two or more risk factor and the sum of those same risk factors.
- Using a combination of terms in xy, x^2y, xy^2 and x^2y^2, we can construct a two factor 2^{nd} order polynomial approximation



- Least squares fit found from 14,641 nodes.

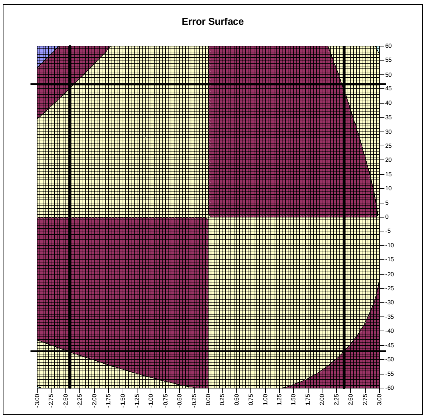




17

Further Considerations
Non Linearity

- Optimal fitting points are given by the intersection of the non-linearity surface and the least squares non linearity approximation function
- No unique solution
- It can be shown that the intersection of the one factor solutions provide *one* solution to the two factor problem
- Same approximation function results from fitting to four nodes as from fitting to 14,641 nodes.



18

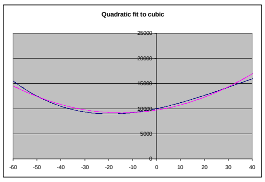
Further Considerations
Non Linearity

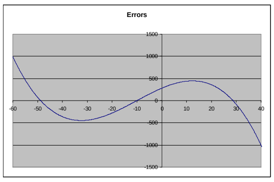
- Can show that the roots to the Legendre polynomials provide optimum nodes single factor polynomials of any order
- Can also show that the intersection of the single factor solutions provide optimum nodes to the following:
  - Two factor 2^{nd} order polynomials
  - Three factor 2^{nd} order polynomials
  - Two factor 3^{rd} order polynomials
- Attempts at a general proof for multifactor polynomials have so far been unsuccessful.

19

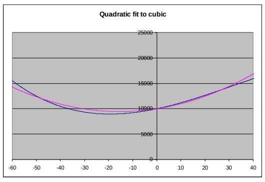
Further Considerations
Constrained solutions

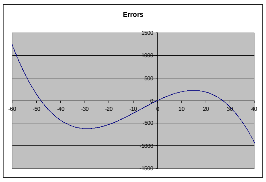
- Consider a quadratic least squares fit over a non-symmetrical domain





- Resulting error at zero may be undesirable
- Can solve directly with constraint that error at zero is nil

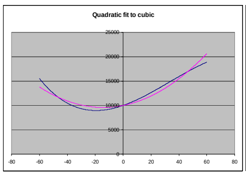


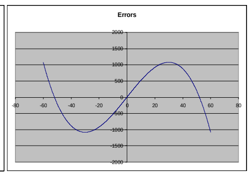


- Resulting fit is optimal given the constraint.

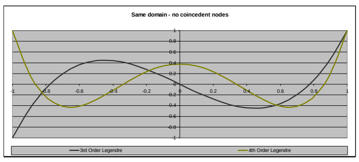
20

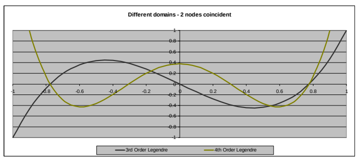
Further Considerations
What about Plan B?





- Minimise the maximum error using Chebyshev nodes
- Adjust the domain to use nested and coincident nodes





- Use a Dampening function
  - Error curve is next term in Legendre series
  - Coefficient is the error at the extreme.

21

Practical Issues

- Stochastic values
  - Approximation error must be minimised
  - Take the average of two results close to, and equidistant from, each node
  - Similarly for each test point
- Error accumulation
  - The errors in marginal risk functions combine to form an error surface
  - These errors may accumulate to exceed materiality limits
  - Lower materiality limits in the marginal risk functions
- Fit to the true non linearity surface
  - Adjust for marginal errors in non linearity before fitting.

22

Outcome

- Efficiency maximised
  - Fit exactly n points for n formula coefficients
- Process is repeatable and objective
  - Based on theories widely accepted and used in engineering, physics and animation
- Less reliance on samples for performance measurement
  - Maximum error is targeted and measured
  - Model limitations can be determined with some confidence
- Better understanding of results
  - Fitting errors are predictable and explainable
- Greater confidence in the internal model results.

23

Questions or comments?

Expressions of individual views by members of The Actuarial Profession and its staff are encouraged.

The views expressed in this presentation are those of the presenter.

24