Institute and Faculty of Actuaries

Proxy models: Lessons from other areas

Peter Murphy, Towers Watson
Dan Messenger, Lloyds Banking Group

10 November 2014

Institute
and Faculty
of Actuaries

Introduction

17 November 2014

Proxy modelling – not just a problem for insurers

Traffic modelling

Climate modelling

Modelling Tsunamis

Defence simulation models

Product design optimisation

Health Economics

Institute and Faculty of Actuaries

Why build proxy models?

17 November 2014

Motivating meta-models

- Simple, low-resolution models are needed for high-level reasoning and communication, decision support, exploratory analysis, and rapidly adaptive calculations.
- Analytical organizations often have large and complex object models, which are regarded as reasonably valid. However, they do not have simpler models and cannot readily develop them by rigorously studying and simplifying the object model.
- Perhaps the object model is hopelessly opaque, the organization no longer has the expertise to delve into the model's innards, or there simply is not enough time to do so. One recourse in such instances is **statistical metamodeling**, which is often referred to as developing a response surface. The idea is to emulate approximately the behavior of the object model with a statistical representation based on a sampling of base-model "data" for a variety of test cases. No deep knowledge of the problem area or the object model is required.

"Motivated Metamodels", Paul Davis & James Bigelow

Motivating meta-models 2

- Modern product and system design typically requires extensive use of simulation based design and analysis codes, such as finite element analysis, computational fluid dynamics, and other computationally intensive models.
- The demand for commercially viable designs and higher performance levels has resulted in an exponential increase in the complexity of these simulation models. Computational demands result in prohibitively high computer times for obtaining results from such complex models, especially in an optimization setting.
- Unreasonably high computer times could also prevent designers from comprehensively exploring the design space, and could ultimately result in underperforming products in the marketplace.
- Metamodeling techniques are widely used in engineering design to address these concerns.

Motivating meta models 3

- Sensitivity analysis is crucial for the best use of the traffic simulation models while also acknowledging that the main obstacle to an extensive use of the most sophisticated techniques is the high number of model runs they usually require.
- To get around this problem, the paper considers the possibility of performing sensitivity analysis not on a model but on its “metamodel”.

Motivating surrogate models

- The origins of many of the surrogate-based optimization techniques in use today can be traced back to geology – more specifically to the science of geostatistics, which has played an important role in mining engineering. Although the applications of geostatistics vary, the fundamental problem is usually formulated as follows.
- The optimum location is sought for a mineral extraction operation – this is usually the maximum ore grade area. The ore grade in a given location can be obtained through drilling a borehole, but this is an expensive operation so it must be comissioned sparingly. The geostatistical solution is to build up a spatial model of ore grade distribution based on the few known borehole values and use the predictions of this model as a guide to identifying the best mining location, or, if the budget permits it, the most informative locations for further boreholes.

Institute and Faculty of Actuaries

How to build proxy models?

17 November 2014

Proxy modelling – standard

Stage 1	Stage 2	Stage 3	Stage 4		Stage 5
Generate Fitting Scenarios	Pick proxy model type	Run models	Determine form of model	Carry out regression	Validation
Expert judgement	Polynomial functions	Standard models	Expert judgement	Least squares	Out of sample testing

Proxy modelling – state of the art?

Stage 1	Stage 2	Stage 3	Stage 4		Stage 5
Generate Fitting Scenarios	Pick proxy model type	Run models	Determine form of model	Carry out regression	Validation
Using SOBOL scenarios	Polynomial Functions	Standard Models	Stepwise algorithms	Least Squares	Cross validation

Proxy modelling – other options

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Generate Fitting Scenarios	Pick proxy model type	Run models	Carry out regression	Validation
Expert Judgement	Polynomial Functions	Standard models	Least Squares	Bootstraping
Random Sampling	Kriging	RSS	Stepwise	Out of sample testing
SOBOL numbers	Neural Networks	2 sim stochastic for LSMC	AIC	Cross validation
Number of runs required	Support Vector Regression		Lasso	

Institute and Faculty of Actuaries

Examples

Climate modelling

- Climate models are exceptionally hungry for computer resources.
- Most of the worlds largest supercomputers are currently running climate projects
- These models can produce up to 5TB of data in a single run
- Researchers are interested in the impact of a large number of input variables on projected global warning.

Proxy modelling – building climate emulators

Stage 1	Stage 2	Stage 3	Stage 4		Stage 5
Generate Fitting Scenarios	Pick proxy model type	Run models	Determine form of model	Carry out regression	Validation
Latin hypercube	Kriging	Standard Models	Stepwise algorithms	Least Squares	Cross validation

Kriging

- Kriging is a form of multi-dimensional interpolation very commonly used in build emulators.
- This is currently seen as best practice for a large number of types of problem.
- The popularity of kriging is due to the fact that computer models are often deterministic (i.e., no random error in the output) and thus interpolating metamodels are desirable.
- Other interpolation like schemes include:
  - Shepherds inverse distance weighted
  - Cubic splines
  - Delaunay interpolation

Kriging - advantages

- All interpolation algorithms estimate the value at a given location as a weighted sum of data values at surrounding locations. Almost all assign weights according to functions that give a decreasing weight with increasing separation distance.
- Kriging assigns weights according to a (moderately) data-driven weighting function, rather than an arbitrary function, but it is still just an interpolation algorithm and will give very similar results to others in many cases. In particular:
  - If the data locations are fairly dense and uniformly distributed throughout the study area, you will get fairly good estimates regardless of interpolation algorithm.
  - If the data locations fall in a few clusters with large gaps in between, you will get unreliable estimates regardless of interpolation algorithm.
  - Almost all interpolation algorithms will underestimate the highs and overestimate the lows; this is inherent to averaging and if an interpolation algorithm didn't average we wouldn't consider it reasonable
- Some advantages of kriging:
  - Helps to compensate for the effects of data clustering, assigning individual points within a cluster less weight than isolated data points (or, treating clusters more like single points)
  - Gives estimate of estimation error (kriging variance), along with estimate of the variable, Z, itself (but error map is basically a scaled version of a map of distance to nearest data point, so not that unique)
  - Availability of estimation error provides basis for stochastic simulation of possible realizations of Z(u)

Kriging

- Assume that the true function $y(x)$, is a realization from a stochastic process

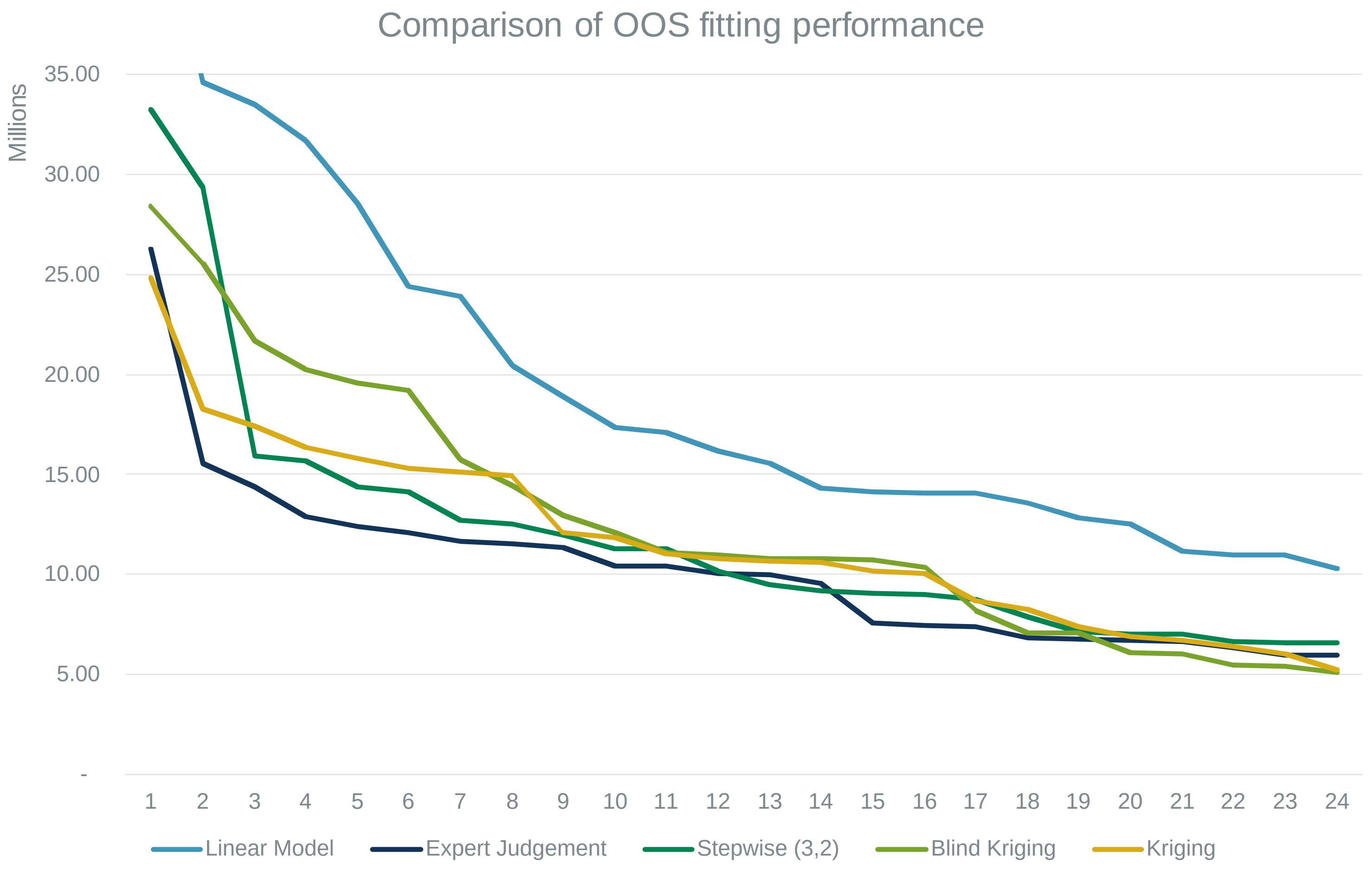
$$Y(x) = \mu(x) + Z(x)$$

- where $\mu(x)$ is the sum of a set of basis functions and $Z(x)$ is a weak stationary stochastic process with mean 0 and covariance function $\sigma^2\psi$.
- The covariance function is defined as cov $\{Y(x+h), Y(x)\} = \sigma^2\psi(h)$, where the correlation function (h) is a positive semi-definite function with $\psi(0) = 1$and $\psi(-h) =$ (h)
- In this formulation $\mu(x)$ is used to capture the known trends, so that $Z(x)$ will be a stationary process. But, in reality, rarely will those trends be known so these terms are sometime omitted.

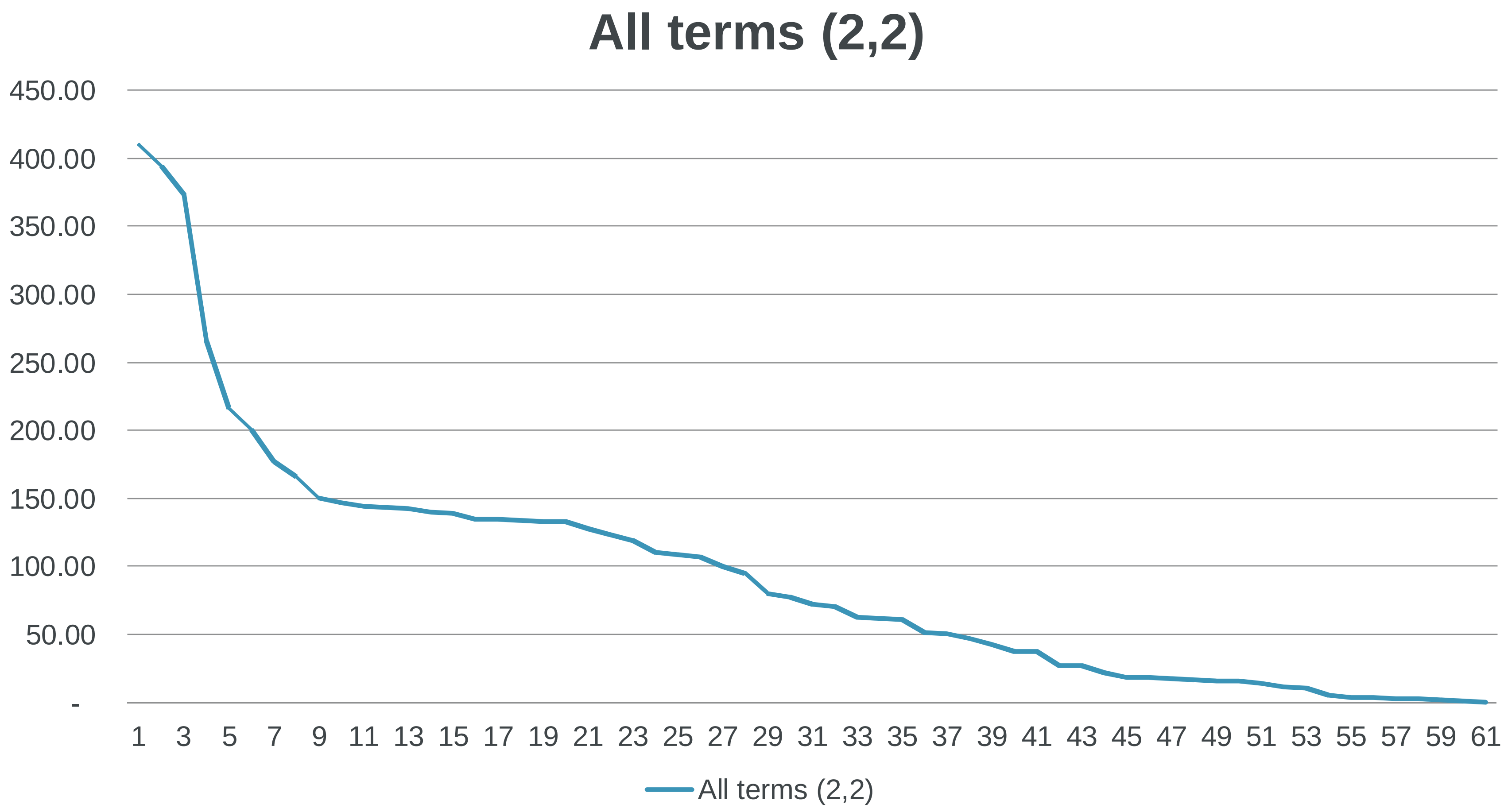
Worked Example

- To test the range of regression approaches we carried out a short test on a sample block of business.
- There was also a range of out of sample test scenarios available.
- There are a number of conclusions
  - The different methods don’t have too material effect.
  - The big problem here is a lack of data (or poor experimental design). This becomes more obvious on detailed investigation of the results, were there are scenarios with very extreme values for some risks on the OOS testing set
  - The linear model isn’t awful
  - Expert Judgment works best, to be expected when there are limitations in the fitting set

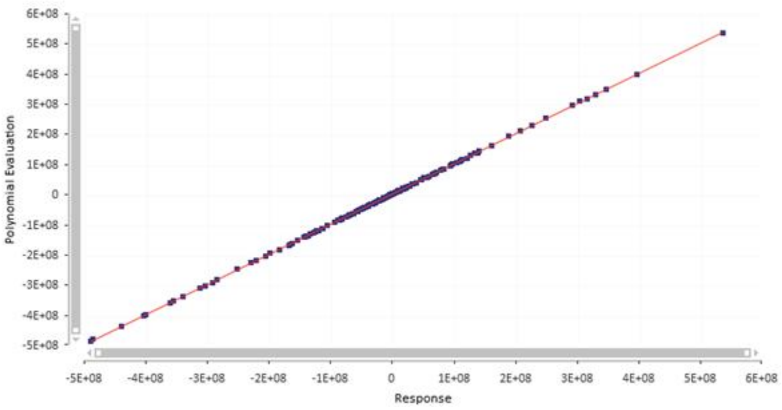
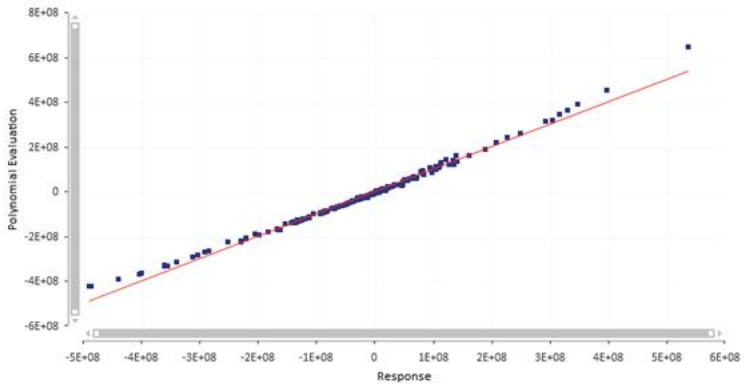
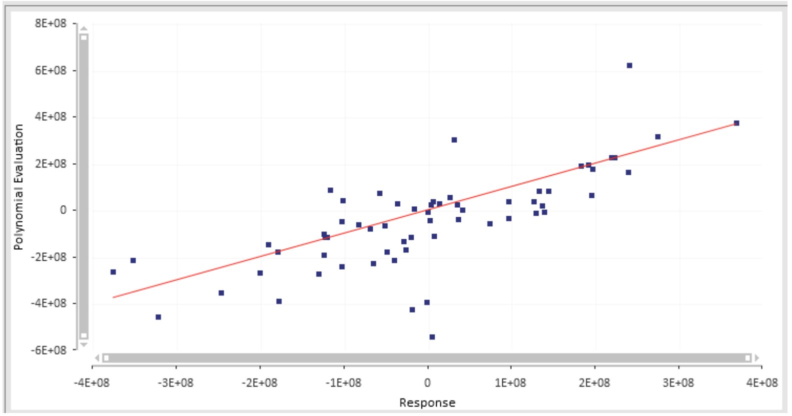
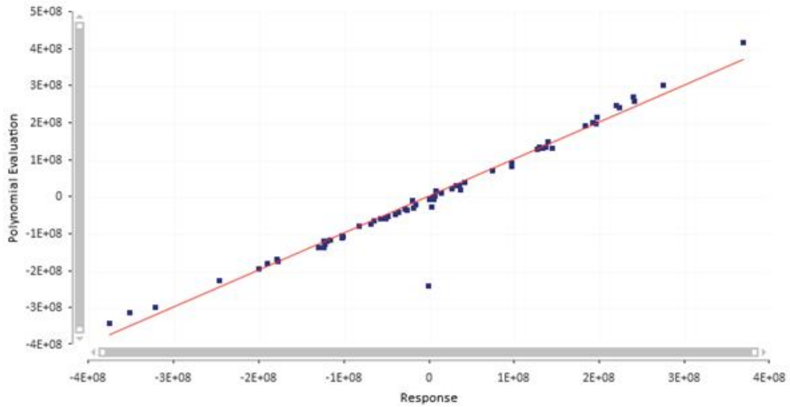
Successful approaches



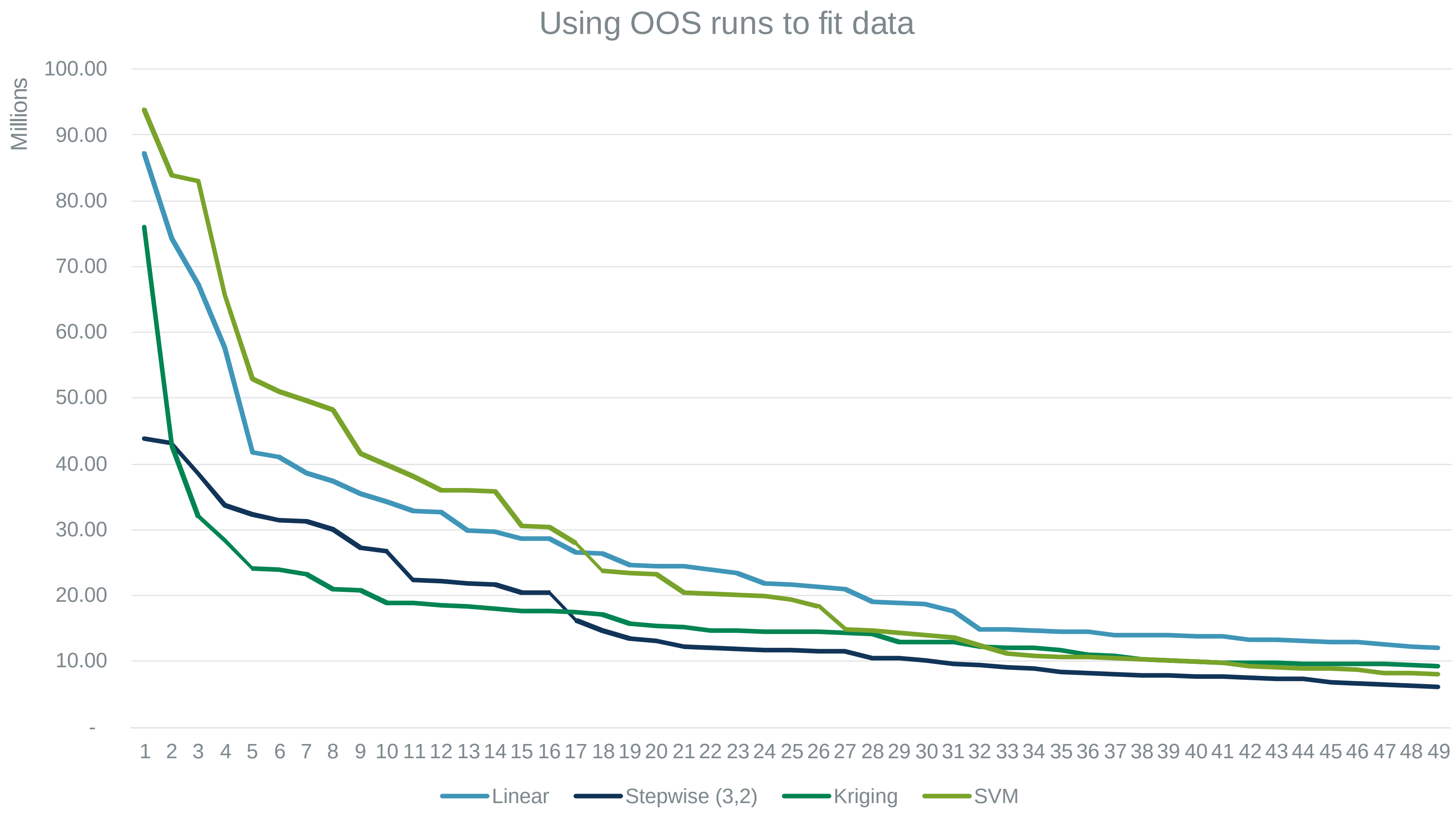
Example OOS errors by fitting type



Comparing linear and quadratic models

	Quadratic	Linear
In Sample		
Out of sample		

We also looked at using the OOS test data for fitting



Other techniques

- There are a range of other techniques that are sometimes used for these types of problems
  - Support Vector Regression
  - Polynomial Chaos Response Surfaces
  - High Dimensional Model representation
- **HDMR** is closely related to the curve fitting methods we use, but with a focus on explaining as much of the variability as possible using the simplest possible models. This seems consistent with some emerging practice in the industry. Trying to use as few cross terms as possible.
- **SVR** is a machine learning approach. Initial investigations have not shown a high level of performance
- **Polynomial Chaos Expansions -** is a way of representing an arbitrary random variable of interest as a function of another random variable with a given distribution, and of representing that function as a polynomial expansion.

Thanks, further reading and references

- Making this talk I have talked to a range of different people to gather background info these included
  - Jeremy Oakley, University of Sheffield
  - Danny Williamson, University of Exeter
- References and useful reading
  - Design and analysis of computer experiments: Jerome Sacks; William J. Welch; Tob. J Mitchell; Henry P Wynn
  - http://www.mucm.ac.uk/Pages/ReadingList.html
- Contact details
  - Danielmessenger@lloydsbanking.com
  - Peter.murphy@towerswatson.com

Questions **Comments**

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

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