Interaction terms in metamodels

# Introduction and motivation

Metamodels are sometimes called surrogate models that can help speed up the calculation of large portfolios of insurance or investment products with complex financial risks. For example, a variable annuity (VA) is an insurance contract which provides complex guarantees that can be difficult to value. In practice, Monte Carlo simulation is commonplace but is computationally intensive. Metamodels have been studied in the literature, e.g., Hejazi and Jackson (2016), Gan and Valdez (2018), Xu, et al. (2018).

The scope of this work is to investigate interaction terms in a Generalized Linear Model (GLM) framework as a metamodel for valuing the complex financial guarantees associated with VA contracts. The aims are how to choose statistically significant interaction terms to the model, how to evaluate performance as predictive model, and how to interpret the resulting effect of the addition of interaction terms.

The nature of this work is expected to provide a cross-road among many disciplines: actuarial science, financial mathematics, statistics, and data science.

# What are the expected outcomes?

Gan (2018) introduced the idea of how to learn from interaction terms in a normal linear regression problem. Within this simple framework, pairwise interactions were examined and evaluated using overlapped group-lasso techniques. The primary expected output from this project is to extend this work of adding interactions of feature variables within a GLM framework. A synthetic dataset with 190,000 VA contracts will be provided for calibrating models; see Gan and Valdez (2017) for description of this dataset.

The whole arena of metamodels for valuing large portfolios of investment and insurance portfolios is relatively new in the literature; there is increasing interest in practice. The outcome of this work will be a useful addition to existing literature. The team is expected to deliver a comprehensive data analysis and report which will outline the work demonstrating the skills used to solve the problem posed.

* A brief discussion of metamodels and their usefulness in the valuation of large portfolios of VA contracts.
* A preliminary data analysis of the synthetic dataset with a focus on the fair market value of the VA guarantees as the response which has a skewed distribution, and possible preliminary investigation of evidence of the presence of interaction terms of available features in the dataset.
* Examine the use of group-lasso or overlapped group-lasso for choosing and learning significant interaction terms. The team should investigate alternative approaches when available, and if the execution is not demanding,.
* Computing and statistical analysis in R is recommended, although use of python is acceptable.
* Analyze the performance of the resulting predictive models using various validation statistics. In metamodeling, a small subset of the entire portfolio is used for calibration, but the entire portfolio is used for performance evaluation.
* Interpret, and possibly intuitively justify, the presence of interaction terms.

# What are the theoretical, data and computational aspects of the project?

Theoretical Aspects

* The team will need to have an understanding of the financial risks associated with Variable Annuity (VA) products.
* The team will need to have an understanding of metamodeling and its usefulness for the valuation of large VA portfolios.
* The team is also expected to have an understanding of some aspect of statistical learning approaches that include but not limited to GLM and regularization (lasso, group-lasso, overlapped group lasso).
* Some actuarial concepts may be necessary for interpretation and more meaningful use of models.

Data and Computational

* The synthetic dataset to be used for analysis will be accessible from a website to be provided. This dataset has been well documented in Gan and Valdez (2017).
* The work is expected to be computationally intensive.
* R is the preferrable language, and the team is expected to be able to navigate through existing packages in order to avoid the additional intensive coding required.

# What are the main challenges?

This project will clearly require good coordination and team work; an excellent team leader may be necessary in order to effectively coordinate the work. All team members may have to learn some concepts within a small time frame. The team can be broken down into smaller groups, although sharing of information is necessary for better coordination.

# Bibliography

S.A. Hejazi and K.R. Jackson (2016) A neural network approach to efficient valuation of large portfolios of variable annuities. Insurance: Mathematics and Economics, 70: 169-181.

G. Gan and E.A. Valdez (2018) Regression modeling for the valuation of large variable annuity portfolios. North American Actuarial Journal, 22(1): 40-54.

W. Xu, Y. Chen, C. Coleman, and T.F. Coleman (2018) Moment matching machine learning methods for risk management of large variable annuity portfolios. Journal of Economic Dynamics and Control, 87: 1-20.

G. Gan (2018) Valuation of large variable annuity portfolios using linear models with interactions. Risks, 6, 71, 1-19.

G. Gan and E.A. Valdez (2017) Valuation of large variable annuity portfolios: Monte Carlo simulation and synthetic datasets. Dependence Modeling, 5: 354-374.