Insurance
Data
Science

Scenario selection with Lasso regression for the valuation of variable annuity portfolio

Presenter: Hang Nguyen

Authors: Hang Nguyen, Michael Sherris, Andres Villegas, Jonathan Ziveyi

20 November 2020

Presentation outline

Motivation

- Variable annuity is an insurance product that couples equity market investment with some form of guaranteed return. Its guaranteed return can be categorized into guaranteed minimum death benefit (GMDB) and guaranteed minimum living benefit (GMLB).
- According to Morningstar, VA net assets reached an all-time high of 2.018 trillion at the end of 2019, a 3.7% increase from $1.946 trillion at the end of 2018.

Motivation

- In the literature on the valuation of a portfolio of VA contracts, a metamodelling approach is often used, which involves:
  - Selecting a sample,
  - Training a metamodel on the sample,
  - Predicting the value of the contracts.
- Existing literture mostly focus on the sampling/ clustering method or the metamodels.
- A large sample will normally improve the accuracy of the predictions but also increase computational load. In our paper, we aim to address this issue.

Literature - Existing approach

1. Select a set G of N_R representative policies from the portfolio $\mathbb{P}$, $G \subset \mathbb{P}$. The representative policies are denoted by p_i, $i = 1, \ldots, N_R$ where $N_R < N$.
2. Calculate the fair market values of the representative contracts by running Monte Carlo simulations:

$$Y(p_i) = \frac{1}{M} \sum_{j=1}^{M} Y_j(p_i), i = 1, \ldots, N_R.$$

Literature - Existing approach

3. Train a regression model (that is, the metamodel) where fair market values of the representative contracts $Y(p_i), i = 1, \ldots, N_R$ are the target variable and policy information of these contracts $X(p_i), i = 1, \ldots, N_R$ are the multivariate inputs. In other words, we train a model to estimate $\hat{f}$ such that $Y(p_i) \approx \hat{f}(X(p_i)), i = 1, \ldots, N_R$.
4. Use the metamodel in Step 3 to estimate the fair market values of the remaining contracts in the portfolio: $\hat{Y}(P_i) = \hat{f}(X(P_i)), i = 1, \ldots, N$.

Literature - Existing approach

- Some Clustering/ Sampling methods to select representative policies:
  - k-prototype algorithm (Gan 2013; Gan and Lin 2015)
  - Latin hypercube sampling method (Gan 2015), conditional Latin hypercube sampling method (Gan and Lin 2017)
  - Truncated fuzzy c-means (Gan and Huang 2017)
  - Transfer learning (Cheng et al. 2019b, 2019a)

Literature - Existing approach

- Some metamodels:
  - Kriging (Gan 2013, 2015; Gan and Lin 2015, 2017; Gan and Huang 2017; Cheng et al. 2019b, 2019b)
  - Spatial interpolation techniques: Kriging, Inverse Distance Weighting(IDW) and Radial Basis Function (RBF) (Hejazi, Jackson, and Gan 2017)
  - Neural network framework: Doyle and Groendyke (2019); Hejazi and Jackson (2016)
  - Regression tree, Neural network and Random forest (Xu et al. 2018); bias-corrected bagging (Gweon, Li, and Mamon 2020)
  - GB2 model (Gan and Valdez 2018) to address skewness in data
  - Linear model with interactions (Gan 2018)

Proposed approach

- Extend the set of representative policies. Use Lasso regression to choose scenarios, reduce computational load.
- The algorithm:

1. Select a set G_1 of representative policies from the portfolio $\mathbb{P}$, $G_1 \subset \mathbb{P}$. Let p_i, $i = 1, \ldots, N_1$, $N_1 < N$ be the policies in G_1.
2. Calculate the fair market values of the representative contracts in G_1 by running Monte Carlo simulations:

$$Y(p_i) = \frac{1}{M} \sum_{j=1}^{M} Y_j(p_i), i = 1, \ldots, N_1.$$

Proposed approach (new steps)

3. New steps to expand the set of representative contracts:

3.1. Select a second group G_2 of representative policies such that $G_1 \subset \mathbb{P}$, $G_2 \subset \mathbb{P}$ and $G_1 \cap G_2 = \emptyset$. Denote $p_{N_1+1}, p_{N_1+2}, \cdots, p_{N_1+N_2}$ the representative policies in G_2, $N_1 + N_2 < N$. The combined set $G = G_1 \cup G_2$ is the set of all representative contracts from G_1 anf G_2.

Proposed approach (new steps)

3.2. Perform a lasso regression on G_1 in order to predict the fair market values $Y(p_i)$ based on the scenario values $Y_1(p_i), Y_2(p_i), \dots, Y_M(p_i)$ for $i = 1, \dots, N_1$. The Lasso regression is formulated by fitting the linear model:

$$Y(p_i) \approx \sum_{j=1}^{M} \beta_j Y_j(p_i),$$

so that it minimizes the cost function:

$$\sum_{i=1}^{N_1} \left(Y(p_i) - \sum_{j=1}^{M} \beta_j Y_j(p_i) \right)^2 + \lambda \sum_{j=1}^{M} \left| \beta_j \right|,$$

Proposed approach (new steps)

3.3. Suppose $W_{j_1}, W_{j_2}, \dots, W_{j_m}$, $m < M$ are the scenarios with non-zero coefficients in the resulting lasso regression model, then we refit a linear regression on G_1 to find a linear function g such that:

$$\begin{aligned} Y(p_i) &\approx g(Y_{j_1}(p_i), Y_{j_2}(p_i)), \dots, Y_{j_m}(p_i))) \\ &= \beta_{j_1} Y_{j_1}(p_i) + \beta_{j_2} Y_{j_2}(p_i) + \dots + \beta_{j_m} Y_{j_m}(p_i), \text{ for } i = 1, \dots, N_1 \end{aligned}$$

Proposed approach (new steps)

3.4. Calculate the values of all contracts in G_2 at the scenarios W_{j_1}, W_{j_2}, ..., W_{j_m} selected in Step (3.2): $Y_{j_1}(p_i)$, $Y_{j_2}(p_i)$, ..., $Y_{j_m}(p_i)$ for $i = N_1 + 1$, $N_1 + 2$, ..., $N_1 + N_2$.

3.5. Use the linear model in Step (3.3), estimate the fair market values of the contracts in G_2 using the values of the contracts at m representative scenarios:

$$\tilde{Y}(p_i) = g(Y_{j_1}(p_i), Y_{j_2}(p_i)), \ldots, Y_{j_m}(p_i))), i = N_1 + 1, \ldots, N_1 + N_2$$

Proposed approach

4. Train a regression model (i.e, the metamodel) on the set $G = G_1 \cup G_2$ of representative contracts, where the target variables are the fair market values of the representative contracts (or the estimated fair market values for contracts in G_2), denoted as $\tilde{Y}(p_i), i = 1, \ldots, N_1 + N_2$. The input variables of the model are the policy information of these contracts $X(p_i), i = 1, \ldots, N_1 + N_2$. In other words, we train a model to estimate $\hat{f}$ such that $\tilde{Y}(p_i) \approx \hat{f}(X(p_i)), i = 1, \ldots, N_1 + N_2$.

5. Use the metamodel in Step (4) to estimate the fair market values of the remaining contracts in the portfolio: $\hat{Y}(P_i) = \hat{f}(X(P_i)), i = 1, \ldots, N$.

Results

- Simulated data (Gan and Valdez 2017): 170,000 policies, 17 riders.
- In the simulated data, Monte Carlo simulation runs 1000 scenarios of the index fund and the forward rate over 30 years, at monthly time step.
- Goal: calculate fair market values of all contracts in the portfolio

Results

- Sampling method: Conditional Latin Hypercube Sampling (cLHS)
- Metamodel: neural network (15 nodes in hidden layer); linear model with interactions
- Input for the metamodel and the sampling method of selecting representative policies, we use the following policy information (that is, $X(p_i)$) as input: Fundvalue (10 fields for 10 funds), riderFee, rollUpRate, gbAmt, gmwbBalance, wbWithdrawalRate, withdrawal, age, timeTMatur, timeIF, productType and gender.
- Output of the metamodel: fair market value of the contracts

Results

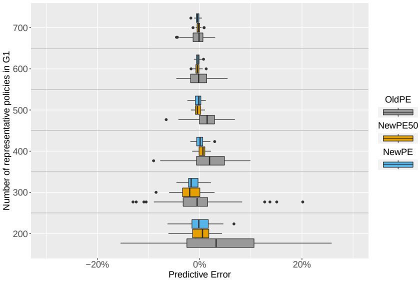

Figure 1: Boxplot comparing the predictive errors of the existing approach and the new approach using neural network regression. The number of additional policies is the twice the original number of representative policies $N_2 = 2N_1$. The boxplot uses the errors from 50 different iterations of fitting the neural network on the same set of representative policies.

Results

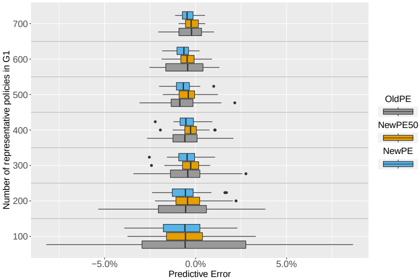

Figure 2: Boxplot comparing the predictive errors of the existing approach and the new approach using linear model with interactions. The number of additional policies is the twice the original number of representative policies $N_2 = 2N_1$. The boxplot uses the errors from fitting the linear model with interactions on 50 different selections of representative policies in G_1.

Other results: inividual contract level

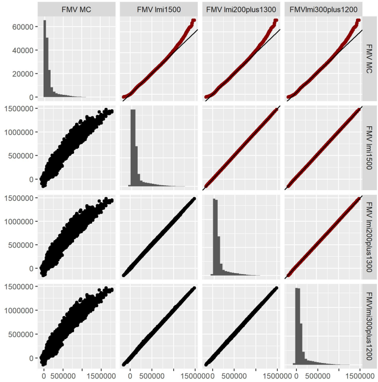

Figure 3: Matrix containing scatter plots (lower triangle) and QQ plots (upper triangle) of the FMV predicted by LMI using one of three approaches: existing approach with 1,500 policies; proposed approach with 200 policies in G_1 and 1,300 policies in G_2; and proposed approach with 300 policies in G_1 and 1,200 policies in G_2.

Other results: inividual contract level

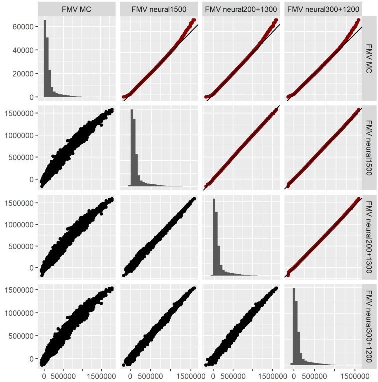

Figure 4: Matrix containing scatter plots (lower triangle) and QQ plots (upper triangle) of the FMV predicted by NN using one of three approaches: existing approach with 1,500 policies; proposed approach with 200 policies in G_1 and 1,300 policies in G_2; and proposed approach with 300 policies in G_1 and 1,200 policies in G_2.

Conclusions

- We propose a method that uses a linear combination of the representative scenarios selected with Lasso regression to estimate the fair market values of the representative contracts.
- By eliminating the need to run the full Monte Carlo simulations to calculate the fair market value of every single representative policy, our approach results in a significant reduction in runtime, which in turn allows us to further extend the set of representative policies and subsequently leads to improvement in accuracy and stability.

Reference I

Cheng, Xiaojuan, Wei Luo B, Guojun Gan, and Gang Li. 2019a. "Deep Neighbor Embedding for Evaluation of Large Portfolios of Variable Annuities." In *Knowledge Science, Engineering and Management. KSEM 2019. Lecture Notes in Computer Science*, 11775:472–80. Springer International Publishing. https://doi.org/10.1007/978-3-030-29551-6.

———. 2019b. "Fast Valuation of Large Portfolios of Variable Annuities via Transfer Learning." In *Trends in Artificial Intelligence. PRICAI 2019. Lecture Notes in Computer Science*, 11672:716–28. Springer International Publishing. https://doi.org/10.1007/978-3-030-29894-4.

Doyle, Daniel, and Chris Groendyke. 2019. "Using Neural Networks to Price and Hedge Variable Annuity Guarantees." *Risks*, no. 7. https://doi.org/10.3390/risks7010001.

Reference II

Gan, Guojun. 2013. "Application of Data Clustering and Machine Learning in Variable Annuity Valuation." *Insurance: Mathematics and Economics* 53 (3). Elsevier B.V.: 795–801. https://doi.org/10.1016/j.insmatheco.2013.09.021.

———. 2015. "Application of Metamodeling to the Valuation of Large Variable Annuity Portfolios." *Paper Presented at Winter Simulation Conference, Huntington Beach, CA, USA*. IEEE, 1103–14. https://doi.org/10.1109/WSC.2015.7408237.

———. 2018. "Valuation of Large Variable Annuity Portfolios Using Linear Models with Interactions." *Risks*. https://doi.org/10.3390/risks6030071.

Reference III

Gan, Guojun, and Jimmy Xiangji Huang. 2017. "A Data Mining Framework for Valuing Large Portfolios of Variable Annuities." *Paper Presented at the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1467–75. https://doi.org/10.1145/3097983.3098013.

Gan, Guojun, and X Sheldon Lin. 2015. "Valuation of Large Variable Annuity Portfolios Under Nested Simulation : A Functional Data Approach." *Insurance: Mathematics and Economics* 62. Elsevier B.V.: 138–50. https://doi.org/10.1016/j.insmatheco.2015.02.007.

Reference IV

Gan, Guojun, and X. Sheldon Lin. 2017. "Efficient Greek Calculation of Variable Annuity Portfolios for Dynamic Hedging: A Two-Level Metamodeling Approach." *North American Actuarial Journal* 21 (2). Taylor & Francis: 161–77. https://doi.org/10.1080/10920277.2016.1245623.

Gan, Guojun, and Emiliano A Valdez. 2017. "Valuation of Large Variable Annuity Portfolios : Monte Carlo Simulation and Synthetic Datasets," 354–74. https://doi.org/10.1515/demo-2017-0021.

———. 2018. "Regression Modeling for the Valuation of Large Variable Annuity Portfolios." *North American Actuarial Journal* 22: 40–54. https://doi.org/10.1080/10920277.2017.1366863.

Reference V

Gweon, Hyukjun, Shu Li, and Rogemar Mamon. 2020. "An Effective Bias-Corrected Bagging Method for the Valuation of Large Variable Annuity Portfolios." *ASTIN Bulletin* 50 (3): 853–71. https://doi.org/10.1017/asb.2020.28.

Hejazi, Seyed Amir, and Kenneth R Jackson. 2016. "A Neural Network Approach to Efficient Valuation of Large Portfolios of Variable Annuities." *Insurance: Mathematics and Economics* 70. Elsevier B.V.: 169–81. https://doi.org/10.1016/j.insmatheco.2016.06.013.

Hejazi, Seyed Amir, Kenneth R Jackson, and Guojun Gan. 2017. "A Spatial Interpolation Framework for Efficient Valuation of Large Portfolios of Variable Annuities." *Quantitative Finance and Economics* 1 (June): 125–44.

Reference VI

Xu, Wei, Yuehuan Chen, Conrad Coleman, and Thomas F Coleman. 2018. "Moment Matching Machine Learning Methods for Risk Management of Large Variable Annuity Portfolios." *Journal of Economic Dynamics and Control* 87 (71771175). Elsevier B.V.: 1–20. https://doi.org/10.1016/j.jedc.2017.11.002.