Shapley decomposition based selection of representative contracts for portfolio valuation of variable annuities

Presenter: Gayani Thalagoda

Joint work with Katja Hanewald, Andrés Villegas, Jonathan Ziveyi
University of New South Wales, Sydney

31st Colloquium on Pensions and Retirement Research
UNSW Sydney, Australia
5-6 December 2023

Overview

Background and motivation

- A variable annuity (VA) is an attractive retirement income product: equity participation & downside protection
- Embedded financial guarantees expose VA insurer's liability to significant market risks
- Regulators encourage frequent, market consistent valuations
  - VM 21 (NAIC, USA): intra-day monitoring
  - APS 117 (APRA, Australia): detailed stochastic calculations
  - Solvency standard (IPSA, NZ): validation on a regular basis
  - IFRS17: high level of granularity to qualify for hedge accounting purposes
- Managing portfolio risks is a major challenge
- In practice, VA insurers use Monte Carlo simulations: challenging & time-consuming (Gan and Valdez, 2017)

Motivational example

- Suppose an insurer plans to rebalance the hedge portfolio at 1:00 pm
- Partial dollar Deltas must be calculated based on the levels of the equity indices at 1:00 pm
- Calculation must be completed within a very short time interval
- Otherwise, the calculated partial dollar Deltas may be very different from the ones at when the calculation is completed
- For a large VA portfolio, it is problematic

Motivation

Under certain conditions, the regulator permits group-level modeling

- Grouping of contracts shall be reflective of the quantity being measured

  AG 43/VM-21: Requirements for PBR of VA

- Aggregation must reflect perceived risk profile

  APS 117 (Australia)

- Grouping requires pooling of similar risks

  IFRS 17 Insurance contracts and Level of Aggregation

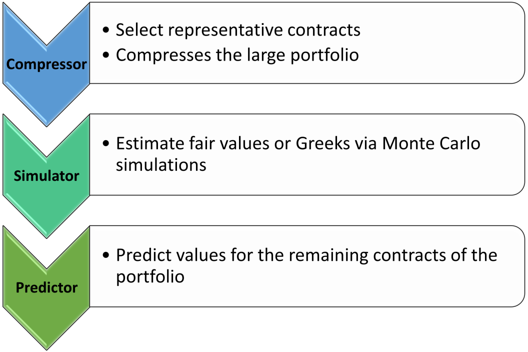
- Group-based valuation models must go through internal model approval process

  Solvency II

Key question

How to determine an optimal grouping?

Existing solution: Meta-modeling framework



Issue: Compression doesn't guarantee the most informative (optimal) sample for the prediction

Literature Gap: Practical implications related to the existing approach

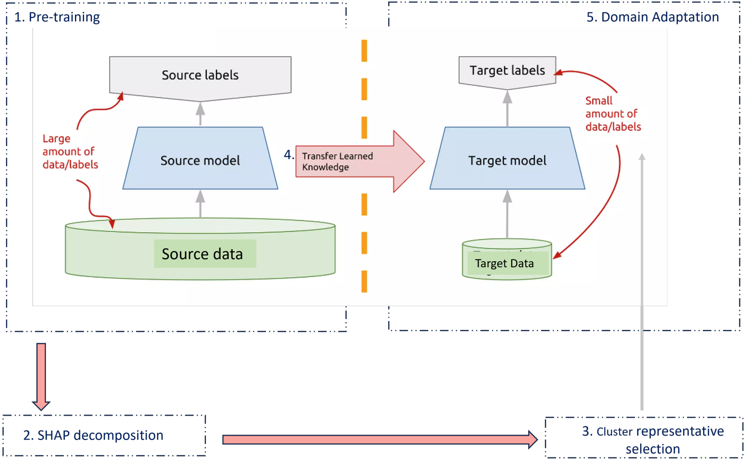
1. Contract selection is independent of quantity being measured $\rightarrow$ lessen explainability
2. Inefficient selection of representative contracts $\rightarrow$ worsen the valuation accuracy

Implication: Less appealing for principle based reserving (PBR) in practice

Solution: Propose a method to select the representative sample in a supervised manner

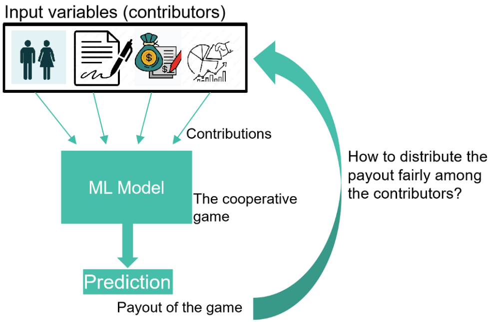
How: By pre-processing the data with SHAP values (*SHapley Additive exPlanations*)

Proposed framework



Shapley framework

- Introduced by Lloyd Shapley in 1953
- Assigning payouts to players based on their contribution to the total payout



Game theory	Model explanation
Game	Prediction task
Players	Input features
Payout	Actual prediction

KernelSHAP explanation

- Lundberg et al. (2017) showed that if we define an explanation model for an instance z' ;

$$g(z') = \psi_0 + \sum_{j=1}^{D} \psi_j z'_j \ ,$$

- train the linear model g by optimizing the following loss function L:

$$L(f, g, \pi_x) = \sum_{z' \in Z} [f(h_x(z')) - g(z')]^2 \pi_x(z')$$

- where the Kernel is;

$$\pi_x(z') = \frac{(D-1)}{\binom{D}{|z'|} |z'| (D - |z'|)} \text{ and}$$

- D is the maximum number of input features, h_x is a mask vector of '0' and '1' for feature presence and $|z'|$ is the number of present features in instance z' ,
- then, the estimated coefficients of the model, the ψ_j's, are the Shapley values.

Transformation of data into SHAP values

	gender	productType	age	ttm	gbAmt
25975	1	7	48.865753	23.764384	72410.947650
16447	0	15	37.438356	8.920548	491185.353584
6780	0	3	44.282192	21.931507	208997.126350
27488	0	8	39.194521	20.600000	133165.250016
6194	0	3	57.367123	20.095890	226772.199973

⇩

	gender	productType	age	ttm	gbAmt
25975	0.568497	0.771683	0.591527	0.676048	0.871734
16447	0.661418	0.700432	0.509260	0.445696	0.642982
6780	0.592795	0.711911	0.546090	0.687851	0.810440
27488	0.568497	0.764130	0.463817	0.644408	0.845290
6194	0.586468	0.712214	0.723272	0.670264	0.801186

The implementation of the decomposition method requires pre-training a neural network on various contract specifications under various market conditions and using it for generalization

Step 1: Pre-training neural network

- Minimize the loss
$$\theta^* = \arg\min_{\theta \in \Theta} \mathbb{E}_{(X^S, Y^S) \sim P^S} [Y^S - f(X^S, \theta)]^2$$
- Let data from the source domain be $\mathcal{D}_S = \{x_i, y_i\}^S = \{\Phi(P_i)^s, V(P_i)^s\}_{i=1}^{N^*}$
- Then the empirical risk minimization becomes;
$$\theta^* = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i \in \mathcal{D}_S} [y_i{}^S - f(x_i{}^S, \theta)]^2$$

Step 2: Obtaining the Shapley decomposition

- Obtaining the SHAP decomposition for each contract using the trained model
- The goal is to decompose the predicted value of the quantity being measured as a sum of marginal attributions from input features

$$\hat{V}(P_i)^s = f(\Phi(P_i)^s, \theta^*) \approx \psi_0 + \sum_{j=1}^{d} \psi_j \tag{1}$$

Step 3 - 4: Selecting the representative sample and running Monte Carlo

3. Clustering the Shapley additive representations to select the representative contracts
4. With the configurations of the target market condition $\mathcal{D}_T$, run the high resolution Monte Carlo simulation to estimate the quantity of interest

$$V(p_i)^T = \frac{1}{M}\sum_{m=1}^{M} V(\Phi(p_i)^T, m),\ i = 1, ..., n \tag{2}$$

Step 5- 6: Fine-tuning and Prediction

5. Fine-tune the transferred knowledge from $(\Phi(P)^S, V^S, p_S)$ to the $(\Phi(P)^T, V^T, p_T)$ following (Yolsinki et al., 2014)

   Fine-tune the model using emprirical risk minimization

$$\theta_T^* = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i \in \mathcal{D}_T} [{y_i}^T - f({x_i}^T, \theta)]^2$$

6. Use the fine-tuned model $f^T(\cdot)$ to predict the value of the remaining contracts

Dataset

- Synthetic dataset of variable annuities constructed by Gan and Valdez (2018)
- 38,000 synthetic VA contracts described by 34 features

Variable	Description
Gender	Gender of the policyholder
Age	Age of the policyholder
Product Type	Product type of the VA policy
GMWB Balance	Guaranteed minimum withdrawal benefit (GMWB) balance
GB Amount	Guaranteed benefit amount
Fund Value i	Account value of the i^{th} fund, for $i = 1, 2, ..., n$
Time to Maturity	Time to maturity in years

Table: https://www2.math.uconn.edu/~gan/software.html

Numerical results I

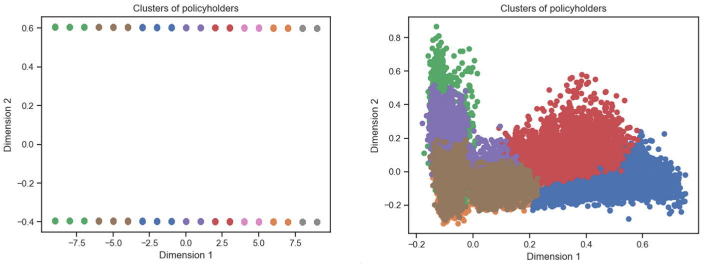


Figure: Formation of clusters (i) Raw input feature space with (ii) SHAP decomposition input space

- Proposed method uses how the features contributed to the prediction as the basis for clustering: effective handles categorical features
- Resulting clusters are explainable

Numerical results II

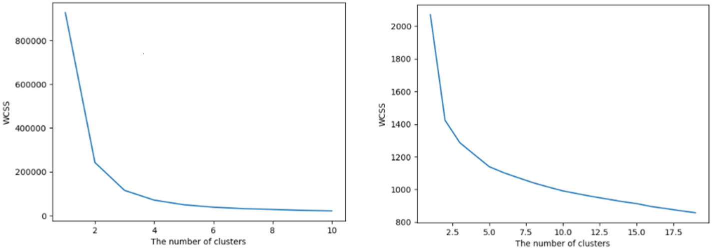


Figure: Comparison WCSS values between clustering from raw values and clustering from SHAP values.

- Within-cluster sum of squares (WCSS) for SHAP value based clustering is lower
- Resulting clusters are more compact $\rightarrow$ better cluster quality

Numerical results III

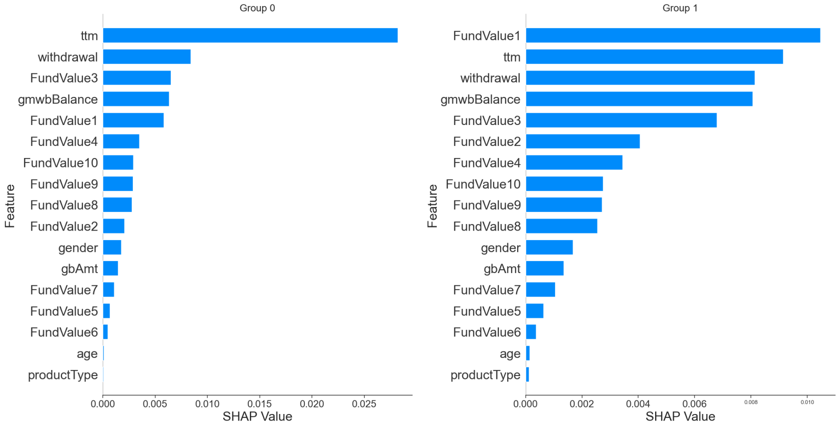


Figure: Graphical illustration of different features contribute differently to the target quantity computed

Numerical results IV

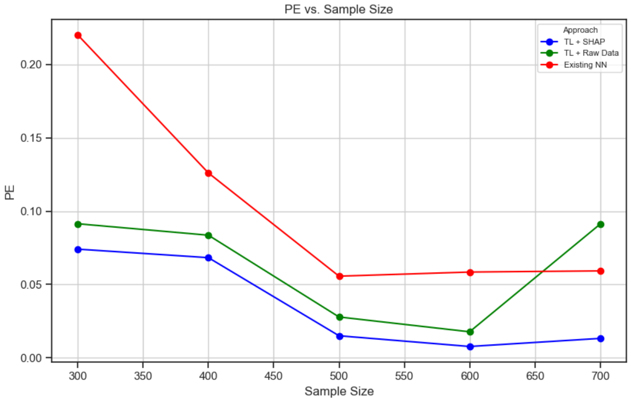


Figure: Portfolio error (PE) across different sample sizes

Numerical results IV

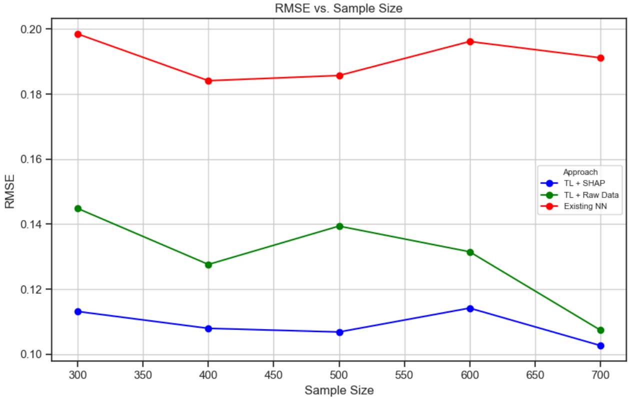


Figure: Room mean sqaured error (RMSE) across different sample sizes

Summary and contributions

- Propose a novel framework to select representatives in an informative manner
- Show that pre-processing with SHAP values
  1. overcomes possible **artificial creation of clusters**
  2. results in **tighter clusters**
  3. **facilitates explainability** for cluster formation
  4. results in **improved PE and RMSE** measures

Thank you

Gayani Thalagoda
g.thalagoda@student.unsw.edu

References I

Feng, R., Goujun, G., & Zhang, N. (2022). Variable annuity pricing, valuation, and risk management: A survey. *Scandinavian Actuarial Journal*, *2022*(10), 867–900. https://doi.org/10.1080/03461238.2022.2049635

Feng, R., Jing, X., & Dhaene, J. (2017). Comonotonic approximations of risk measures for variable annuity guaranteed benefits with dynamic policyholder behavior. *Journal of Computational and Applied Mathematics*, *311*, 272–292.

Gan, G. (2022). Metamodeling for variable annuity valuation: 10 years beyond kriging. *2022 Winter Simulation Conference (WSC)*, 915–926. https://doi.org/10.1109/WSC57314.2022.10015284

Gan, G., & Valdez, E. A. (2017). Valuation of large variable annuity portfolios: Monte carlo simulation and synthetic datasets. *Dependence Modeling*, *5*(1), 354–374.

Gan, G., & Valdez, E. A. (2018). Nested stochastic valuation of large variable annuity portfolios: Monte carlo simulation and synthetic datasets. *Data*, *3*(3), 31.

References II

- Godin, F., Hamel, E., Gaillardetz, P., & Ng, E. H.-M. (2023).Risk allocation through shapley decompositions, with applications to variable annuities. *ASTIN Bulletin: The Journal of the IAA*, *53*(2), 311–331.
- Lin, X., & Yang, S. (2020).Efficient dynamic hedging for large variable annuity portfolios with multiple underlying assets. *ASTIN Bulletin: The Journal of the IAA*, *50*(3), 913–957.
- Lundberg, S. M., & Lee, S.-I. (2017).A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.
- Richman, R., & Wüthrich, M. V. (2023).Conditional expectation network for shap. *arXiv preprint arXiv:2307.10654*.
- Rosenstein, M. T., Marx, Z., Kaelbling, L. P., & Dietterich, T. G. (2005).To transfer or not to transfer. *NIPS 2005 workshop on transfer learning*, *898*(3).
- Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016).A survey of transfer learning. *Journal of Big data*, *3*(1), 1–40.

References III

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014).How transferable are features in deep neural networks? *Advances in neural information processing systems, 27.*

Numerical Results: additional

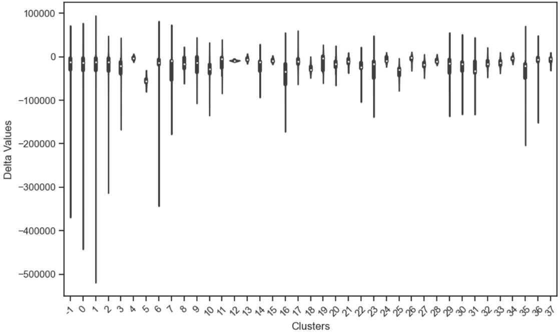


Figure: Formation of clusters with raw values

- Clustering with SHAP values generate notably heterogeneous clusters and use its centroids as model points might not be enough