North American Actuarial Journal

ISSN: 1092-0277 (Print) 2325-0453 (Online) Journal homepage: https://www.tandfonline.com/loi/uaaj20

Valuation of Large Variable Annuity Portfolios with Rank Order Kriging

Guojun Gan & Emiliano A. Valdez

To cite this article: Guojun Gan & Emiliano A. Valdez (2020) Valuation of Large Variable Annuity Portfolios with Rank Order Kriging, North American Actuarial Journal, 24:1, 100-117, DOI: 10.1080/10920277.2019.1617169

To link to this article: https://doi.org/10.1080/10920277.2019.1617169

Published online: 06 Sep 2019.

North American Actuarial Journal, 24(1), 100–117, 2020

ISSN: 1092-0277 print / 2325-0453 online
DOI: 10.1080/10920277.2019.1617169

Valuation of Large Variable Annuity Portfolios with Rank Order Kriging

Guojun Gan and Emiliano A. Valdez
Department of Mathematics, University of Connecticut, Storrs, Connecticut

Metamodels, which simplify the simulation models used in the valuation of large variable annuity portfolios, have recently increased in popularity. The ordinary kriging and the GB2 (generalized beta of the second kind) regression models are examples of metamodels used to predict fair market values of variable annuity guarantees. It is well known that the distribution of fair market values is highly skewed. Ordinary kriging does not fit skewed data well but depends on only a few parameters that can be estimated straightforwardly. GB2 regression can handle skewed data but parameter estimation can be quite challenging. In this article, we explore the rank order kriging method, which can handle highly skewed data and depends only on a single parameter, for the valuation of large variable annuity portfolios. Our numerical results demonstrate that the rank order kriging method performs remarkably well in terms of fitting the skewed distribution and producing accurate estimates of fair market values at the portfolio level.

1. INTRODUCTION

A variable annuity (VA) refers to an attractive insurance product that provides upside participation and downside protection in both bull and bear markets. A main feature of variable annuities is that they contain guarantees (Hardy 2003; Ledlie et al. 2008; The Geneva Association 2013). Due to the attractive guarantee features, lots of variable annuity contracts have been sold in the past decade. According to the Insured Retirement Institute, for example, new sales in the United States in 2016 were 102 billion dollars and the total assets under management was 1.9 trillion.

Due to limited reinsurance capacity, dynamic hedging is widely adopted by insurance companies to mitigate the financial risks arising from their VA business. In fact, the hedging programs helped to save the industry about $40 billion in September and October 2008 during the global financial crisis (Chopra et al. 2009). However, dynamic hedging requires computing Greeks or sensitivities of the guarantees to major market factors. Because VA guarantees are relatively complex, their fair market values, which are used to calculate Greeks, cannot be evaluated explicitly. Insurance companies rely heavily on Monte Carlo simulation to calculate the fair market values. One major drawback of Monte Carlo simulation is that it is computationally intensive to value a large VA portfolio because every VA contract needs to be projected over many scenarios for a long time horizon (Dardis 2016).

During the past few years, metamodeling approaches have been used to address the computational problem arising from VA hedging. A metamodel is a surrogate model intended to simplify the original model. Using a metamodeling approach to estimate the fair market value (FMV) of a portfolio of VA contracts involves four major steps (Barton 2015):

1. Select a small number of representative VA contracts from the portfolio (experimental design).
2. Run Monte Carlo simulation to calculate the FMVs of the selected VA contracts.
3. Build a metamodel based on the selected VA contracts and the corresponding FMVs (predictive modeling).
4. Use the metamodel to estimate the FMVs of all VA contracts in the portfolio.

As seen from the above steps, metamodeling techniques do not require valuing the full portfolio. Because the metamodel is much more computationally efficient than the Monte Carlo simulation model, it substantially reduces the valuation time.

Address correspondence to Guojun Gan, Department of Mathematics, University of Connecticut, 341 Mansfield Road, Storrs, CT 06268-1009. E-mail: guojun.gan@uconn.edu

The metamodeling approach depends on two major components: an experimental design and a predictive model. Research to date has focused on these two components. Gan (2013) used the *k*-prototypes clustering algorithm for the experimental design and ordinary kriging as the metamodel. Gan and Lin (2015) used the *k*-prototypes algorithm to select representative VA contracts and universal kriging for functional data as the metamodel. Because the *k*-prototypes algorithm is not efficient for selecting a moderate number (e.g., 200) of representative VA contracts, Latin hypercube sampling (LHS; Gan 2015) and the truncated fuzzy *c*-means algorithm (Gan and Huang 2017) have been proposed to select representative contracts. In particular, Gan and Valdez (2016) investigated several experimental design methods and found that clustering and LHS are comparable in accuracy and are better than other methods such as random sampling. Gan and Lin (2017) used the LHS and conditional LHS methods with ordinary and universal kriging to calculate dollar deltas quickly for daily hedging purpose.

Hejazi and Jackson (2016) and Xu et al. (2018) used neural networks as metamodels for the valuation of large VA portfolio. Hejazi, Jackson and Gan (2017) treated the valuation of large VA portfolios as a spatial interpolation problem and studied several interpolation methods, including inverse distance weighting and radial basis function. Gan and Valdez (2017a) studied the use of a copula to model the dependency of partial dollar deltas and found that using a copula does not improve the prediction accuracy of the metamodel because the dependency is well captured by the covariates. Gan and Valdez (2018) proposed the use of the GB2 (generalized beta of the second kind) distribution to model the fair market values in order to address the skewness typically observed in the distribution of the fair market values. Gan (2018) studied the use of linear models with interaction terms as metamodels for the valuation of large VA portfolios.

Among the aforementioned metamodels, ordinary kriging and the GB2 regression model are two popular choices. The advantage of ordinary kringing over the GB2 regression model is that the former depends only on a few parameters that can be estimated straightforwardly. However, one drawback of ordinary kriging is that the dependent variable (i.e., the fair market value) is assumed to follow a Gaussian distribution. This assumption is not appropriate for the fair market value of the guarantees as the distribution of fair market values is positively skewed. Although the GB2 regression model addresses the skewness of the dependent variable, estimating the parameters of the GB2 regression model posed additional challenges (Gan and Valdez 2018).

In this article, we study the use of rank order kriging (Pachepsky, Radcliffe, and Selim 2003) to model the fair market values of VA guarantees. Rank order kriging is also known as quantile kriging. This method estimates the rank order of the dependent variable given the values of the independent variables and transforms the estimated rank order to the original scale. Rank order kriging has the advantages that it depends only on a few parameters and that it is effective for handling highly skewed data (Juang, Lee, and Ellsworth 2000).

The remainder of this article is organized as follows. In Section 2, we give a description of the data we use to demonstrate the usefulness of rank order kriging for VA valuation. In Section 3, we introduce and describe the rank order kriging method in detail. In Section 4, we present some numerical results to show the performance of rank order kriging. Section 5 concludes the article with some remarks.

2. DESCRIPTION OF THE DATA

We use the synthetic data set produced in Gan and Valdez (2017b) to evaluate the performance of the rank order kriging method. This data set consists of a portfolio of 190,000 VA policies and their fair market values. Each VA policy is characterized by 45 features. Some features (e.g., fund fees) have identical values for all policies. We exclude these features and use the remaining features as explanatory variables. The following list contains the variables we use for modeling the fair market values:

- gender: Gender of the policyholder
- productType: Product type of the VA contract
- gmwbBalance: Guaranteed Minimum Withdrawal Benefit (GMWB) balance
- gbAmt: Guaranteed benefit amount
- FundValue$i$: Account value of the ith fund, for $i = 1, 2, \ldots, 10$
- age: Age of the policyholder. It is calculated from the birth date and the valuation date.
- ttm: Time to maturity in years. It is calculated from the valuation date and the maturity date.

Among the variables listed above, gender and productType are the only categorical variables. The portfolio contains 76,007 (about 40%) policies with female policyholders. There are 19 product types and each product type has 10,000 policies.

TABLE 1
Summary Statistics of the Continuous Variables

	Min.	1st Quarter	Median	Mean	3rd Quarter	Max.
gmwbBalance	0.00	0.00	0.00	35,611.54	0.00	499,708.73
gbAmt	0.00	186,864.95	316,225.98	326,834.59	445,940.63	1,105,731.57
FundValue1	0.00	0.00	12,635.17	33,433.87	49,764.15	1,099,204.71
FundValue2	0.00	0.00	15,107.17	38,542.81	56,882.55	1,136,895.87
FundValue3	0.00	0.00	10,043.96	26,740.18	39,199.69	752,945.34
FundValue4	0.00	0.00	10,383.79	26,141.80	39,519.79	610,579.68
FundValue5	0.00	0.00	9,221.26	23,026.50	35,023.00	498,479.36
FundValue6	0.00	0.00	13,881.41	35,575.67	52,981.06	1,091,155.87
FundValue7	0.00	0.00	11,541.47	29,973.25	44,465.70	834,253.63
FundValue8	0.00	0.00	11,931.41	30,212.11	45,681.16	725,744.64
FundValue9	0.00	0.00	11,562.79	29,958.29	44,302.35	927,513.49
FundValue10	0.00	0.00	11,850.05	29,862.24	44,967.78	785,978.60
age	34.52	42.03	49.45	49.49	56.96	64.46
ttm	0.59	10.34	14.51	14.54	18.76	28.52

TABLE 2
Summary Statistics of the Fair Market Values

	Min.	1st Quarter	Median	Mean	3rd Quarter	Max.
fmv	−94.94	−5.14	12.49	67.59	66.81	1,536.70

Note: Numbers are in thousands.

Table 1 shows some summary statistics of the continuous variables. From the table, we see that except for age and ttm, all variables are dollar amounts and have zeros. Table 2 shows some summary statistics of the fair market values. Many policies have negative fair market values because for these policies the benefit is less than the risk charge. From Table 2 we also see that the mean is much higher than the median, indicating that the distribution of the fair market values is positively skewed. The histogram shown in Figure 1 confirms the skewness of the distribution.

The fair market values shown above were calculated by Monte Carlo simulation, which is computationally intensive. As reported in Gan and Valdez (2017b), it would take a single CPU (central processing unit) about 108 hours to calculate the fair market values for all of the policies in the portfolio.

3. RANK ORDER KRIGING

In this section, we describe the rank order kriging method in detail. Figure 2 shows a high-level procedure of this method. From the figure, we see that in the rank order kriging method, standardized ranks are estimated and back transformed to get the fair market values in the original scale.

3.1. Ordinary Kriging for Ranks

To describe the ordinary kriging method for ranks, let $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k$ be the k representative VA policies and let $v_1, v_2, \ldots, v_k$ be the corresponding fair market values (or other quantities of interest). For $j = 1, 2, \ldots, k$, let u_j be the standardized rank of v_j; that is,

$$u_j = u(\mathbf{z}_j) = \frac{r(v_j)}{k}, \tag{1}$$

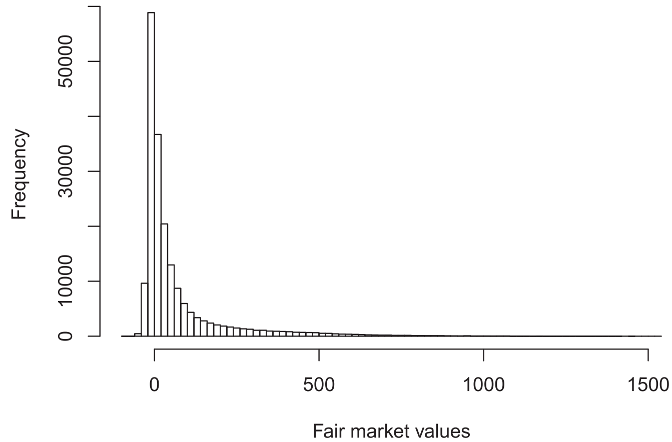


FIGURE 1. A Histogram of the Fair Market Values. *Note:* Numbers in the horizontal axis are in thousands.

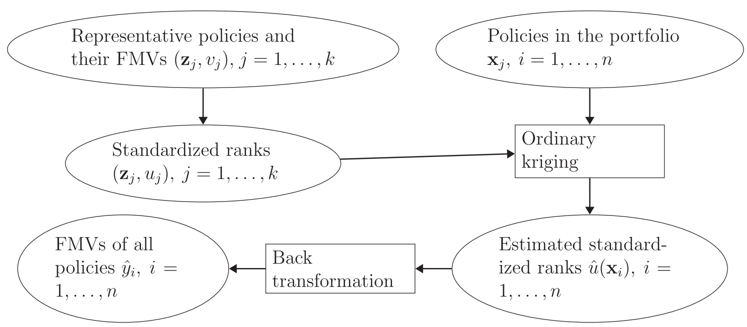


FIGURE 2. Sketch of the Rank Order Kriging Method.

where $r(v_j) \in \{1, 2, ..., k\}$ is the rank order of v_j; that is, $r(v_j)$ is the position of v_j when $v_1, v_2, ..., v_k$ are arranged in ascending order. The standardized rank transformation is monotonically increasing.

Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ be the VA policies in a portfolio, where n is the number of VA policies in the portfolio. Under an ordinary kriging model, the standardized rank of the fair market value of the guarantees embedded in the ith policy $\mathbf{x}_i$ is assumed to be (Cressie 1993; Juang, Lee, and Ellsworth 2000):

$$U(\mathbf{x}_i) = \mu + \delta(\mathbf{x}_i), \quad i = 1, 2, ..., n,$$

where μ is an unknown constant and $\delta(\cdot)$ is a zero-mean intrinsically stationary spatial process. In this model, $U(\mathbf{x}_i)$ can be predicted as (Cressie 1993)

$$\hat{u}(\mathbf{x}_i) = \sum_{j=1}^{m} w_{ij} u(\mathbf{z}_j), \tag{2}$$

where $u(\mathbf{z}_j)$ is the standardized rank order of the jth representative policy as defined in Equation (1) and $w_{i1}, w_{i2}, ..., w_{im}$ are the kriging weights. These kriging weights are obtained by solving the following linear equation system:

$$\begin{pmatrix} V_{11} & \cdots & V_{1m} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ V_{m1} & \cdots & V_{mm} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} w_{i1} \\ \vdots \\ w_{im} \\ \theta_i \end{pmatrix} = \begin{pmatrix} D_{i1} \\ \vdots \\ D_{im} \\ 1 \end{pmatrix}, \tag{3}$$

where θ_i is the Lagrange multiplier to ensure the sum of the kriging weights equal to one and V_{ls} and D_{il} are semivariograms that describe the degree of spatial dependence of the fair market values. Mathematically, V_{ls} and D_{il} are calculated as

$$V_{ls} = \gamma(||\mathbf{z}_l - \mathbf{z}_s||), \tag{4a}$$

$$D_{il} = \gamma(||\mathbf{x}_i - \mathbf{z}_l||), \tag{4b}$$

where $||\cdot||$ is the L^2 norm (i.e., the Euclidean distance) and $\gamma(h)$ is a semivariogram function defined as (Chiles and Delfiner 2012):

$$\gamma(h) = \frac{1}{2}\mathrm{Var}[u(\mathbf{z}_1) - u(\mathbf{z}_2)], \quad h = ||\mathbf{z}_1 - \mathbf{z}_2||.$$

There are several theoretical semivariogram models, including linear, spherical, exponential, and Gaussian models. For example, the linear semivariogram model can be specified as (Chiles and Delfiner 2012)

$$\gamma(h) = bh,$$

where b is a parameter. A linear model is appropriate when spatial variability increases linearly with distance. The exponential semivariogram model can be specified as (Isaaks and Srivastava 1990)

$$\gamma(h) = 1 - \exp\left(-\frac{3}{\beta}h\right), \quad l, s = 1, 2, \ldots, m,$$

where $\beta > 0$ is a parameter.

To select a suitable theoretical semivariogram model, we investigate the empirical semivariogram given by (Juang, Lee, and Ellsworth 2000)

$$\gamma_e(h) = \frac{1}{2|S(h)|} \sum_{(\mathbf{s},\mathbf{t}) \in S(h)} [u(\mathbf{s}) - u(\mathbf{t})]^2, \tag{5}$$

where $S(h)$ is a set of all pairs of policies that have a distance of h; that is,

$$S(h) = \{(\mathbf{s}, \mathbf{t}) : ||\mathbf{s} - \mathbf{t}|| = h, \mathbf{s} \in X_0, \mathbf{t} \in X_0\},$$

and $|S(h)|$ is the number of elements in $S(h)$. Here $X_0 = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\}$ is the set of representative policies.

3.2. Back Transformation

Most of the rank order kriging estimates $\hat{u}(\mathbf{x}_i)$ are between 0 and 1. There are a few estimates that are outside of this interval. We need to transform these kriging estimates back to the original scale. This can be done by an interpolation model. Before transforming the rank order kriging estimates back to the scale of the original data, we need to correct the smoothing effect of these estimates (Yamamoto 2010).

A uniform random variable on an interval (a, b) has an expected value of $\frac{a+b}{2}$ and a variance of $\frac{(b-a)^2}{12}$. Because the standardized rank order $U(\mathbf{x})$ has a uniform distribution on $(1/n, 1)$, it has a variance of $\frac{1}{12}\left(\frac{n-1}{n}\right)^2$, which is approximately $\frac{1}{12}$ when n is large. However, the empirical variance of the rank order kriging estimates $\hat{u}(\mathbf{x}_1), \ldots, \hat{u}(\mathbf{x}_n)$ is usually less than $\frac{1}{12}$ due to the smoothing effect of the kriging method. Yamamoto (2005) proposed a procedure to correct the smoothing effect caused by kriging. This procedure involves cross-validation, which is time consuming. In this article, we use a different and simple approach to correct the smoothing effect.

In particular, we modify the rank order kriging estimates as follows:

$$u^*(\mathbf{x}_i) = \frac{r(\hat{u}(\mathbf{x}_i))}{n}, \tag{6}$$

where $r(\hat{u}(\mathbf{x}_i))$ is the position of $\hat{u}(\mathbf{x}_i)$ when $\hat{u}(\mathbf{x}_1), \hat{u}(\mathbf{x}_2), ..., \hat{u}(\mathbf{x}_n)$ are arranged in ascending order. Since

$$\{u^*(\mathbf{x}_i) : i = 1, 2, ..., n\} = \left\{\frac{1}{n}, \frac{2}{n}, ..., 1\right\},$$

the empirical variance of $u^*(\mathbf{x}_1), ..., u^*(\mathbf{x}_n)$ is equal to $\frac{1}{12}(\frac{n-1}{n})^2$. In addition, the modified rank order kriging estimates fall within the interval $(0, 1]$, whereas the raw estimates may fall outside the interval.

Then we can transform the modified rank order kriging estimates back to the original scale. This can be done by using a regression model or an interpolation method. For simplicity, we use a linear interpolation method given as follows:

$$y_i^* = v_{i_1} + \frac{u^*(\mathbf{x}_i) - u_{i_1}}{u_{i_2} - u_{i_1}}(v_{i_2} - v_{i_1}), \tag{7}$$

where i_1 and i_2 are indices such that u_{i_1} and u_{i_2} are the ranks closest to $u^*(\mathbf{x}_i)$. The estimated fair market values $y_1^*, y_2^*, ..., y_n^*$ usually contain biases due to the data transformation (Garcia et al. 2010). As a result, we need to adjust the biases in these estimates to get the final estimates of the fair market values. We use the multiplicative bias adjustment method (Garcia et al. 2010) to adjust the biases as follows:

$$\hat{y}_i = \hat{c} y_i^*, \tag{8}$$

where $\hat{c} = \frac{\hat{\mu}}{\mu_m}$ is the ratio of an estimate of the mean of the fair market values over the mean of the fair market values obtained from the kriging model; for example,

$$\hat{\mu} = \frac{1}{k}\sum_{i=1}^{k} v_i, \quad \mu_m = \frac{1}{n}\sum_{i=1}^{n} y_i^*.$$

This back-transformation approach, although ad hoc, is efficient and works well for our purpose as demonstrated by the numerical results in this article.

4. NUMERICAL RESULTS

In this section, we present some numerical results to demonstrate the performance of the rank order kriging method. In particular, we compare the rank order kriging method with the ordinary kriging method and the GB2 regression model.

4.1. Experimental Setup

Metamodeling has two major components: an experimental design method and a metamodel. The experimental design method is used to select representative VA contracts. The metamodel is used to predict the fair market values of all of the VA contracts in the portfolio. In this article, our focus is to develop metamodels. In order to compare the performance of metamodels in a consistent manner, we use the same experimental design method for all metamodels. Gan and Valdez (2016) compared several experiment design methods and found that clustering-based methods produce good results. In our numerical experiments, we use the hierarchical k-means algorithm (Gan and Valdez 2019) to select representative VA policies. In Appendix B we provide numerical results of the metamodels for two additional experimental design methods.

Another important factor to consider in experimental design is the number of representative VA policies. Intuitively, the number of representative VA policies cannot be too large or too small. Using a too large number of representative VA policies will increase the runtime significantly because all representative VA policies have to be valued by Monte Carlo simulation. Using a too small number of representative VA policies is also not appropriate because the resulting metamodel may not produce accurate predictions. In previous studies (e.g., Gan and Lin 2015; Gan 2018), the number of representative VA contracts is determined to be 10 times the number of predictors, including the dummy variables converted from categorical variables. In this article, we follow this strategy to determine the number of representative VA policies. Because there are 34 predictors, we start with 340 representative VA contracts. We also use 170 representative VA contracts to see the impact of the number of representative VA contracts on the performance of the metamodels.

4.2. Validation Measures

To compare the performance of the metamodels in terms of accuracy, we use the following four validation measures: the percentage error, the mean error, the R^2 (Frees 2009), and the concordance correlation coefficient of quantiles (Lin 1989). The first two measures assesses the accuracy of the metamodels at the portfolio level. The third measure assesses the accuracy at the individual policy level. The last measure is used to assess the agreement between two empirical distributions.

Let y_i and $\hat{y}_i$ denote the fair market value of the ith VA policy in the portfolio that is calculated by Monte Carlo simulation and estimated by a metamodel, respectively. Then the percentage error (PE) and the mean error (ME) are defined as follows:

$$PE = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\sum_{i=1}^{n} y_i}, \tag{9}$$

$$ME = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n}, \tag{10}$$

where n is the number of VA policies in the portfolio. The lower the absolute values of the PE and the ME, the better the results.

The R^2 is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{11}$$

where $\bar{y}$ is the average of $y_1, y_2, ..., y_n$. The lower the R^2, the better the results.

The concordance correlation coefficient for quantiles ($CCCQ$) is used to measure the agreement between the quantiles of two samples. To define this measure, we let q_j and $\hat{q}_j$ be the $\frac{100j}{m}$ th quantile of $\{y_1, y_2, ..., y_n\}$ and $\{\hat{y}_1, \hat{y}_2, ..., \hat{y}_n\}$, respectively, for $j = 1, 2, ..., m$, where m is a positive integer (e.g., $m = 1000$). Then $CCCQ$ is defined as follows (Lin 1989):

$$CCCQ = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}, \tag{12}$$

where ρ is the correlation between $(q_1, q_2, ..., q_m)$ and $(\hat{q}_1, \hat{q}_2, ..., \hat{q}_m)$, σ_1 and μ_1 are the standard deviation and the mean of $(q_1, q_2, ..., q_m)$, respectively, and σ_2 and μ_2 are the standard deviation and the mean of $(\hat{q}_1, \hat{q}_2, ..., \hat{q}_m)$, respectively. A higher $CCCQ$ means more agreement between the quantiles of the fair market values obtained from two models.

4.3. Results

We applied the rank order kriging method, the ordinary kriging method, and the GB2 regression model to predict the fair market values with $k = 340$ and $k = 170$ representative VA policies. Table 3 shows the accuracy of the three models when 340 representative VA policies were used. Except for the R^2, all validation measures show that rank order kriging performs the

TABLE 3
Accuracy of the Metamodels When $k = 340$ Representative VA Policies Were Used

	PE	ME	R^2	$CCCQ$
Rank order kriging	0.0018	0.1236	0.8121	0.9958
Ordinary kriging	0.0032	0.2191	0.8009	0.9557
GB2 regression	0.0382	2.5811	0.8227	0.9853

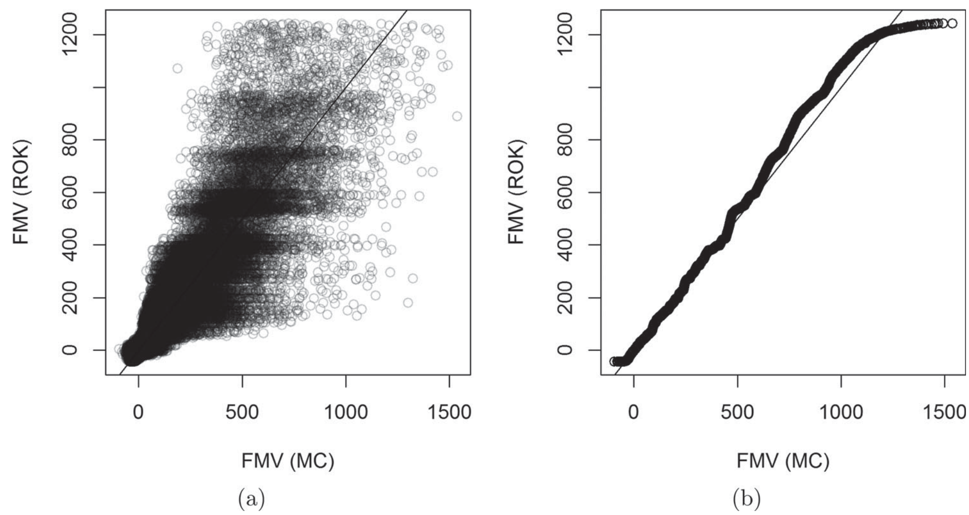


FIGURE 3. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by Rank Order Kriging When $k = 340$ Representative VA Policies Were Used.

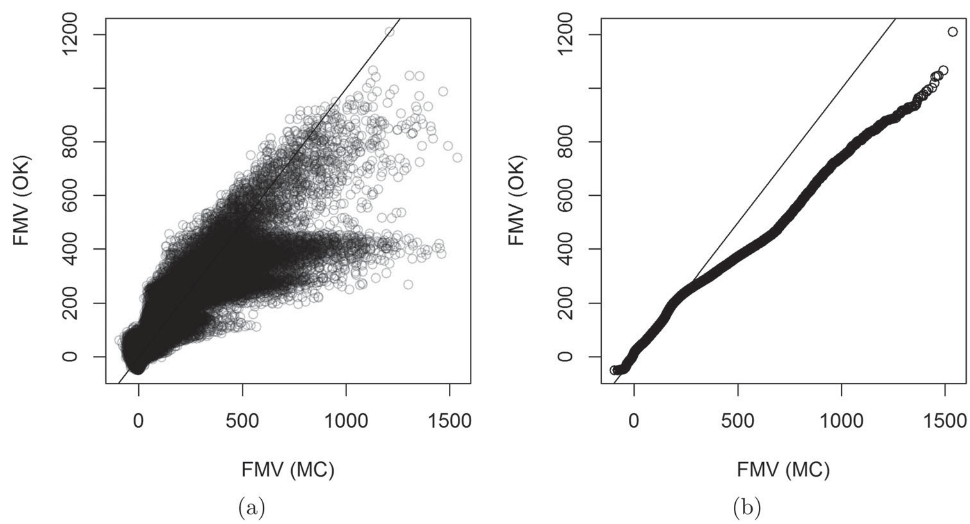


FIGURE 4. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by Ordinary Kriging When $k = 340$ Representative VA Policies Were Used.

best among the three models. For example, the percentage error obtained by rank order kriging is 0.18%, which is much lower than that obtained by the GB2 regression model. In terms of R^2, the GB2 regression model performs the best because it produced the highest R^2.

From Table 3, we also see that rank order kriging produced a *CCCQ* value of 0.9958, which is close to 1. This indicates that the distribution of the fair market values estimated by rank order kriging matches quite well the distribution of the fair market values calculated by Monte Carlo.

Figure 3 shows the scatterplot and the Q–Q plot of the fair market values estimated by rank order kriging and those calculated by Monte Carlo when 340 representative VA policies were used. The scatter plot shows that the estimation at the individual policy level is not very accurate. However, the points scatter quite symmetrically around the 45° line. The Q–Q plot shows that the rank order kriging method did a good job of fitting the skewed data. The points fall closely to the 45° line.

Figures 4 and 5 show the scatter and Q–Q plots obtained from the ordinary kriging method and the GB2 regression model, respectively. The scatterplot in Figure 4(a) shows that the estimates obtained from ordinary kriging are biased. The Q–Q plot in Figure 4(b) shows that ordinary kriging does not fit the tail well. Figure 5(a) shows that the scatterplot produced by GB2

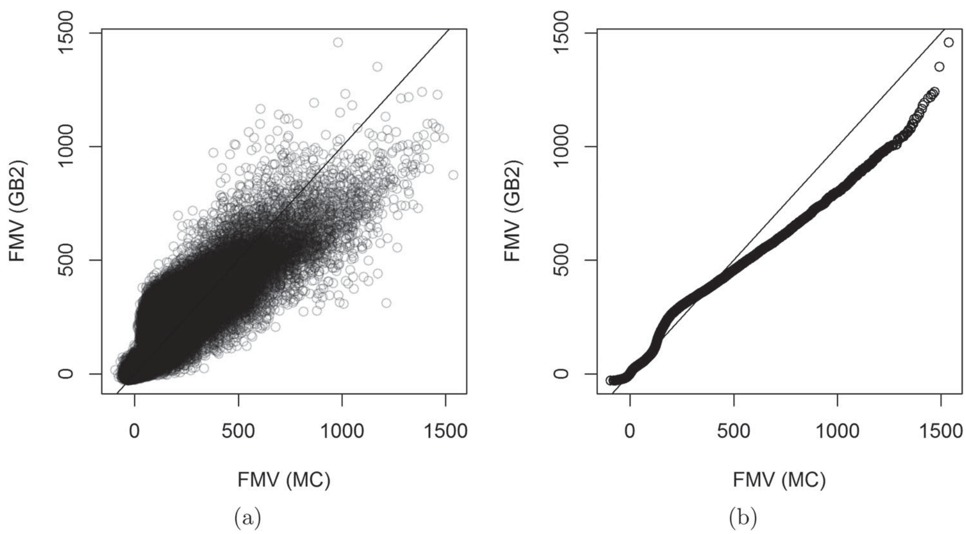


FIGURE 5. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by GB2 Regression When $k = 340$ Representative VA Policies Were Used.

TABLE 4
Summary Statistics of the Fair Market Values Estimated by Different Methods When $k = 340$ Representative VA Policies Were Used

	Min.	1st Quarter	Median	Mean	3rd Quarter	Max.
Monte Carlo	–94.94	–5.14	12.49	67.59	66.81	1,536.70
Rank order kriging	–43.54	–5.02	11.88	67.71	59.90	1,244.11
Ordinary kriging	–50.38	6.97	28.62	67.81	70.67	1,210.32
GB2 regression	–29.24	–4.10	19.68	70.17	57.23	1,459.24

Note: Numbers are in thousands.

regression is quite symmetric around the 45° line. The Q–Q plot in Figure 5(b) shows that GB2 regression fits the tail well although the fit is a little bit off in the middle.

Comparing Figures 3, 4, and 5, we see that the rank order kriging method works pretty well in terms of fitting skewed data. The ordinary kriging method does not fit well the tail of skewed data as expected. Table 4 shows some summary statistics of the fair market values estimated by the three methods as well as those calculated by Monte Carlo. From the table, we see that summary statistics obtained from the rank order kriging match well those obtained from Monte Carlo.

Figures 6(a) and 6(b) shows the empirical semivariograms of rank order kriging and ordinary kriging that are estimated from the data, respectively. From the figures, we see that the empirical semivariogram of ordinary kriging has a wide range. Both empirical semivariograms show that a linear theoretical semivariogram can be used. However, the ordinary kriging method does not work with the linear semivariogram due to the wide range of semivariogram values. Instead, we used the exponential semivariogram for the ordinary kriging method.

Figure 7(a) and (b) shows the histograms of the rank order kriging estimates and the modified rank order kriging estimates, respectively. Figure 7(a) shows that the standardized ranks estimated by rank order kriging have bell shapes and a reduced variance due to the smoothing effect of the kriging method. In fact, the variance of the rank order kriging estimates is around 0.066, which is less than the variance of a uniform variable on $(1/n, 1)$. Figure 7(b) shows that the modified rank order kriging estimates have a uniform distribution.

Table 5 shows the accuracy of the three models when 170 representative VA policies were used. Again, except for the R^2, all validation measures indicate that the rank order kriging method performs the best among the three models. The R^2 values show that the rank order kriging method does not fit well the data at the individual policy level. However, the *CCCQ* value shows that rank order kriging fits the distribution pretty well.

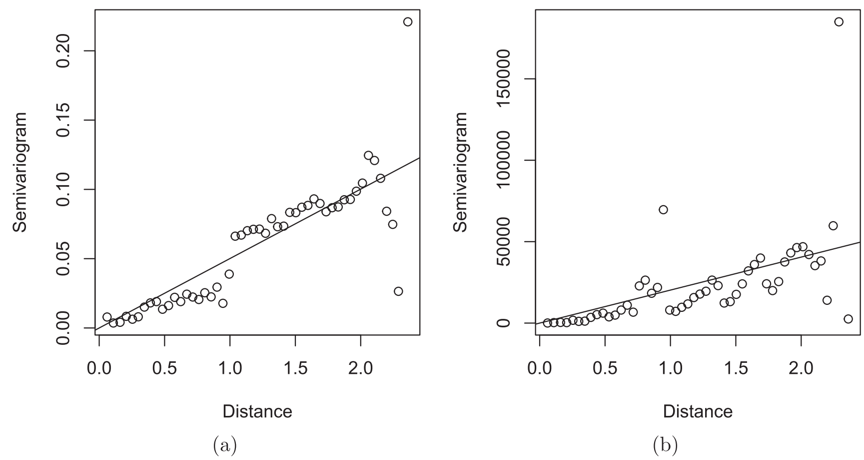


FIGURE 6 (a) Empirical Semivariograms of Rank order Kriging and (b) Ordinary Kriging When $k = 340$ Representative VA Policies Were Used.

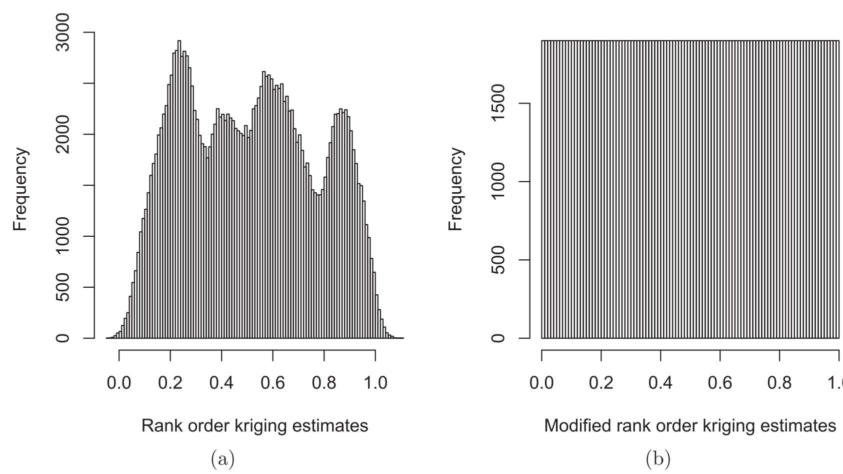


FIGURE 7. (a) Histograms of the Rank Order Kriging Estimates and (b) Modified Rank Order Kriging Estimates When $k = 340$ Representative VA Policies Were Used.

TABLE 5
Accuracy of the Metamodels When $k = 170$ Representative VA Policies Were Used

	PE	*ME*	R^2	*CCCQ*
Rank order kriging	0.0040	0.2729	0.2381	0.9947
Ordinary kriging	0.0865	5.8491	0.6061	0.8568
GB2 regression	–0.0073	–0.4951	0.8097	0.9784

Figures 8, 9, and 10 show the scatter and QQ plots produced by the three models when 170 representative VA policies were used. The QQ plots in Figures 8(b), 9(b), and 10(b) show that the rank order kriging method fit the skewed data quite well. However, the scatterplot in Figure 8(a) shows that the fit produced by rank order kriging is not very accurate at the individual policy level. The rank order kriging method overestimated and underestimated the fair market values of some policies. The number of overestimates is comparable to the number of underestimates. As a result, the scatterplot is quite symmetric around the 45° line. Table 6 shows some summary statistics of the fair market values obtained from the three methods when 170 representative VA policies were used. The numbers in the table show that rank order kriging performed the best in terms of matching the quantiles.

Figure 11 shows the empirical semivariograms of the rank order kriging method and the ordinary kriging method when 170 representative VA policies were used. We see patterns similar to those observed before and the semivariograms can be

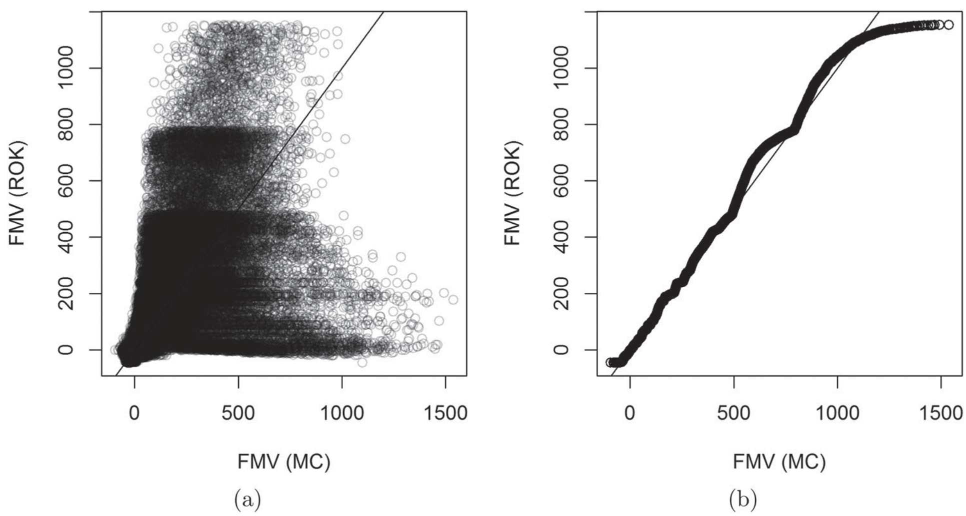


FIGURE 8. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by Rank Order Kriging When $k = 170$ Representative VA Policies Were Used.

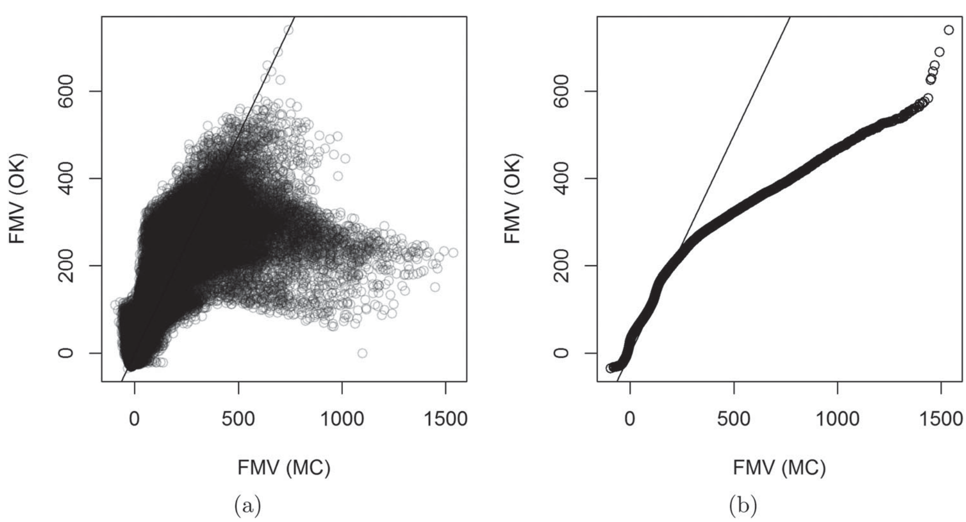


FIGURE 9. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by Ordinary Kriging When $k = 170$ Representative VA Policies Were Used.

modeled linearly. However, the linear semivariogram for ordinary kriging has a wide range of values and does not work for ordinary kriging. Again, we used the exponential semivariogram for ordinary kriging. Figure 12 shows the histograms of the rank order kriging estimates and the modified estimates. We also see patterns similar to those observed before. The distribution of the rank order kriging estimates has a bell shape and a reduced variance.

Finally, Table 7 shows the runtime of the three models. The runtime includes the time used to fit the model and the time used to predict the fair market values. The time used to select representative policies is not included. From the table, we see that the GB2 regression model is the fastest model among the three models. Rank order kriging and ordinary kriging are much slower than GB2 regression due to the large number of distance calculations. However, the runtime used by rank order kriging with $k = 340$ was around 5 minutes, which is much less than the 108 hours used by Monte Carlo simulation (Gan and Valdez 2017b).

In summary, the numerical results show that the rank order kriging method is able to fit quite well the distribution of fair market values that is highly skewed. In addition, the rank order kriging method produces accurate estimates at the portfolio level and its runtime is similar to that of ordinary kriging.

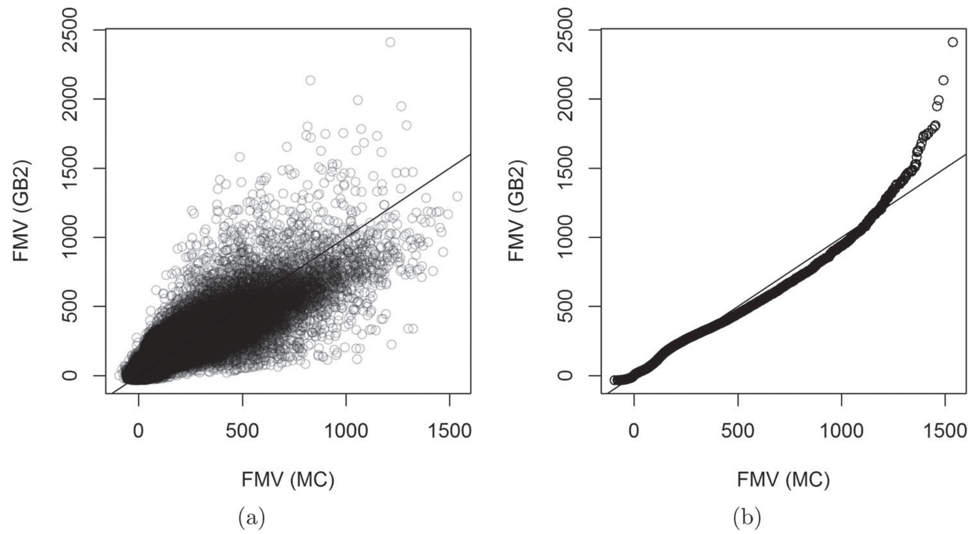


FIGURE 10. Scatter and Q–Q plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by GB2 Regression When $k = 170$ Representative VA Policies Were Used.

TABLE 6
Summary Statistics of the Fair Market Values Estimated by Different Methods When $k = 170$ Representative VA Policies Were Used

	Min.	1st Quarter	Median	Mean	3rd Quarter	Max.
Monte Carlo	–94.94	–5.14	12.49	67.59	66.81	1536.70
Rank order kriging	–93.01	–5.68	11.83	64.63	65.66	1228.46
Ordinary kriging	–68.37	2.12	25.77	68.40	76.07	964.12
GB2 regression	–37.47	–3.56	18.07	66.12	67.27	1098.31

Note: Numbers are in thousands.

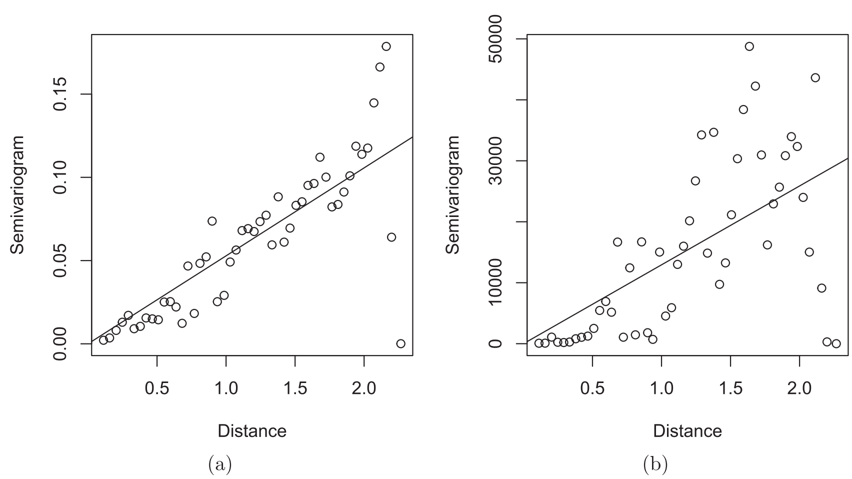


FIGURE 11. (a) Empirical Semivariograms of Rank Order Kriging and (b) Ordinary Kriging When $k = 170$ Representative VA Policies Were Used.

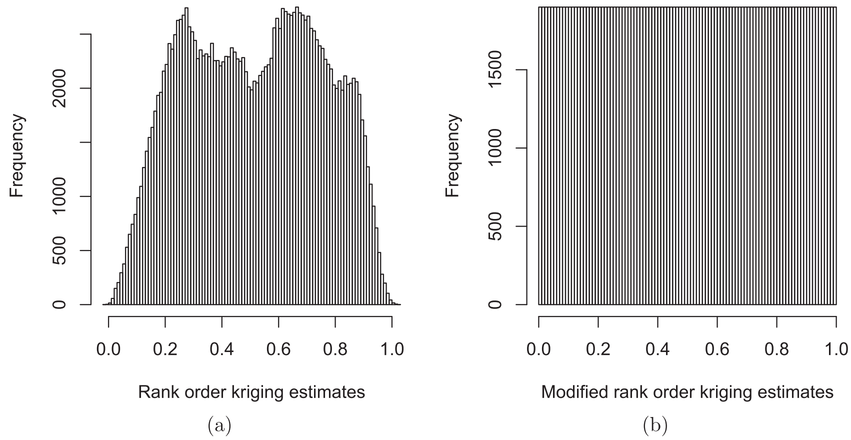


FIGURE 12. (a) Histograms of the Rank Order Kriging Estimates and (b) Modified Rank Order Kriging Estimates When $k=170$ Representative VA Policies Were Used.

TABLE 7
Runtime of the Metamodels

	Rank order kriging	Ordinary kriging	GB2 regression
$k=340$	291.75	286.62	6.38
$k=170$	129.12	129.29	3.33

Note: Numbers are in seconds.

5. CONCLUSIONS

The ordinary kriging and the GB2 regression model are two popular metamodels used to predict the fair market values of VA guarantees. The advantage of ordinary krinigng over the GB2 regression model is that the former is less demanding in terms of parameter estimation. However, ordinary kriging assumes that the dependent variable (i.e., the fair market value) follows a Gaussian distribution. This assumption is not appropriate for the fair market value of the guarantees because the distribution of fair market values is positively skewed. Although the GB2 regression model addresses the skewness of the dependent variable, estimating the parameters of the GB2 regression model is quite challenging (Gan and Valdez 2018).

In this article, we studied the use of rank order kriging to predict the fair market values of VA guarantees. Rank order kriging utilizes data transformation and thus is able to handle highly skewed data. Our numerical results show that the rank order kriging method works well as expected. In particular, we compared rank order kriging with GB2 regression and ordinary kriging and found that rank order kriging is better than GB2 regression and ordinary kriging in terms of fitting highly skewed data.

FUNDING

The authors acknowledge the financial support provided by the Committee on Knowledge Extension Research (CKER) of the Society of Actuaries.

ORCID

Guojun Gan http://orcid.org/0000-0003-3285-7116

REFERENCES

Barton, R. R. 2015. Tutorial: Simulation metamodeling. In *Proceedings of the 2015 Winter Simulation Conference*, 1765–79. Piscataway, NJ: IEEE Press.
Chiles, J.-P., and P. Delfiner. 2012. *Geostatistics: Modeling spatial uncertainty*, 2nd ed. Hoboken, NJ: Wiley.

Chopra, D., O. Erzan, G. de Gantes, L. Grepin, and C. Slawner. 2009. Responding to the variable annuity crisis. McKinsey Working Papers on Risk.
Cressie, N. 1993. *Statistics for spatial data*, Rev. ed. Hoboken, NJ: Wiley.
Cummins, J., G. Dionne, J. B. McDonald, and B. Pritchett. 1990. Applications of the GB2 family of distributions in modeling insurance loss processes. *Insurance: Mathematics and Economics* 9 (4):257–72.
Dardis, T. 2016. Model efficiency in the U.S. life insurance industry. *The Modeling Platform* (3):9–16.
Frees, E. W. 2009. *Regression modeling with actuarial and financial applications*. Cambridge University Press.
Gan, G. 2013. Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics* 53 (3): 795–801.
Gan, G. 2015. Application of metamodeling to the valuation of large variable annuity portfolios. In *Proceedings of the Winter Simulation Conference*, 1103–14. Piscataway, NJ: IEEE Press.
Gan, G. 2018. Valuation of large variable annuity portfolios using linear models with interactions. *Risks* 6 (3):71.
Gan, G., and J. Huang. 2017. A data mining framework for valuing large portfolios of variable annuities. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1467–75. New York, NY: ACM.
Gan, G., and X. S. Lin. 2015. Valuation of large variable annuity portfolios under nested simulation: A functional data approach. *Insurance: Mathematics and Economics* 62:138–50.
Gan, G., and X. S. Lin. 2017. Efficient greek calculation of variable annuity portfolios for dynamic hedging: A two-level metamodeling approach. *North American Actuarial Journal* 21 (2):161–77.
Gan, G., and E. A. Valdez. 2016. An empirical comparison of some experimental designs for the valuation of large variable annuity portfolios. *Dependence Modeling* 4 (1):382–400.
Gan, G., and E. A. Valdez. 2017a. Modeling partial greeks of variable annuities with dependence. *Insurance: Mathematics and Economics* 76: 118–34.
Gan, G., and E. A. Valdez. 2017b. Valuation of large variable annuity portfolios: Monte carlo simulation and synthetic datasets. *Dependence Modeling* 5: 354–74.
Gan, G., and E. A. Valdez. 2018. Regression modeling for the valuation of large variable annuity portfolios. *North American Actuarial Journal* 22 (1): 40–54.
Gan, G., and E. A. Valdez. 2019. Data clustering with actuarial applications. *North American Actuarial Journal*.
Garcia, V. C., K. M. Foley, E. Gego, D. M. Holland, and S. T. Rao. 2010. A comparison of statistical techniques for combining modeled and observed concentrations to create high-resolution ozone air quality surfaces. *Journal of the Air & Waste Management Association* 60 (5):586–95.
The Geneva Association. 2013. Variable annuities—An analysis of financial stability. https://www.genevaassociation.org/sites/default/files/research-topics-document-type/pdf_public/ga2013-variable_annuities_0.pdf (accessed August 21, 2019).
Hardy, M. 2003. *Investment guarantees: Modeling and risk management for equity-linked life insurance*. Hoboken, NJ: John Wiley & Sons.
Hejazi, S. A., and K. R. Jackson. 2016. A neural network approach to efficient valuation of large portfolios of variable annuities. *Insurance: Mathematics and Economics* 70:169–81.
Hejazi, S. A., K. R. Jackson, and G. Gan. 2017. A spatial interpolation framework for efficient valuation of large portfolios of variable annuities. *Quantitative Finance and Economics* 1 (2):125–44.
Isaaks, E., and R. Srivastava. 1990. *An introduction to applied geostatistics*. Oxford, UK: Oxford University Press.
Juang, K.-W., D.-Y. Lee, and T. R. Ellsworth. 2000. Using rank-order geostatistics for spatial interpolation of highly skewed data in a heavy-metal contaminated site. *Journal of Environmental Quality* 30 (3):894–903.
Ledlie, M. C., D. P. Corry, G. S. Finkelstein, A. J. Ritchie, K. Su, and D. C. E. Wilson. 2008. Variable annuities. *British Actuarial Journal* 14 (2): 327–389.
Lin, L. I.-K. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45 (1):255–68.
Pachepsky, Y., D. E. Radcliffe, and H. M. Selim. 2003. *Scaling methods in soil physics*. Boca Raton, FL: CRC Press.
Xu, W., 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.
Yamamoto, J. K. 2005. Correcting the smoothing effect of ordinary kriging estimates. *Mathematical Geology* 37 (1):69–94.
Yamamoto, J. K. 2010. Backtransforming rank order kriging estimates. *Geologia USP: Série Cientifica* 10 (2):101–15.

Discussions on this article can be submitted until October 1, 2020. The authors reserve the right to reply to any discussion. Please see the Instructions for Authors found online at http://www.tandfonline.com/uaaj for submission instructions.

APPENDIX A. GB2 REGRESSION MODEL

The GB2 regression model was proposed to predict the fair market values of VA guarantees (Gan and Valdez 2018). A GB2 random variable has the following probability density function (Cummins et al. 1990):

$$f(x) = \frac{|a|}{bB(p,q)} \left(\frac{x}{b}\right)^{ap-1} \left[1 + \left(\frac{x}{b}\right)^{a}\right]^{-p-q}, \quad x>0, \tag{A.1}$$

where $a \neq 0$, $p > 0$, $q > 0$, $b > 0$, and $B(p, q)$ is the beta function. The parameters a, p, and q are referred to as the shape parameters of the GB2 distribution. The parameter b is called the scale parameter. When $-p < \frac{1}{a} < q$, the expectation exists and is given by

$$E[X] = \frac{bB\left(p + \frac{1}{a}, q - \frac{1}{a}\right)}{B(p, q)}. \tag{A.2}$$

Let Y denote the fair market value of guarantees embedded in a VA policy. Because Y can be negative, the shifted fair market value

$$X = Y + c$$

is modeled with a GB2 distribution, where c is the shift parameter to be estimated from the data.

Independent or regressor variables are incorporated through the scale parameter b in order to make sure that the expectations can be calculated for all policies. The method of maximum likelihood is used to estimate the parameters. Let s be the number of VA policies in the experimental design. For $i = 1, 2, ..., s$, let v_i be the fair market value of the guarantees embedded in the ith VA policy in the experimental design. Then the log-likelihood function of the model is defined as follows:

$$\begin{aligned} L(\boldsymbol{\theta}) = s \ln \frac{|a|}{B(p, q)} - ap \sum_{i=1}^{s} \mathbf{z}_i^T \boldsymbol{\beta} + (ap - 1) \sum_{i=1}^{s} \ln (v_i + c) \\ -(p + q) \sum_{i=1}^{s} \ln \left[1 + \left(\frac{v_i + c}{\exp (\mathbf{z}_i^T \boldsymbol{\beta})} \right)^a \right], \end{aligned} \tag{A.3}$$

where $\boldsymbol{\theta} = (a, p, q, c, \boldsymbol{\beta})$ and $\mathbf{z}_i$ is a numerical vector representing the ith VA policy in the experiment design. It is worth noting that estimating the parameters of the GB2 regression model is quite challenging. Gan and Valdez (2018) proposed a multi-stage procedure to estimate the parameters.

The fitted GB2 regression model is used to predict the fair market values of guarantees for the portfolio. Let n be the number of VA policies in the portfolio and $\mathbf{x}_i$ the numeric vector representing the ith VA policy in the portfolio. Then the fair market value of guarantees for the ith VA policy can be estimated as follows:

$$\hat{y}_i = \frac{\exp (\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) B\left(\hat{p} + \frac{1}{\hat{a}}, \hat{q} - \frac{1}{\hat{a}}\right)}{B(\hat{p}, \hat{q})} - \hat{c}, \quad i = 1, 2, ..., n, \tag{A.4}$$

where $\hat{a}, \hat{p}, \hat{q}, \hat{c}$, and $\hat{\boldsymbol{\beta}}$ are parameters estimated from the data.

APPENDIX B. COMPARISON OF EXPERIMENTAL DESIGNS

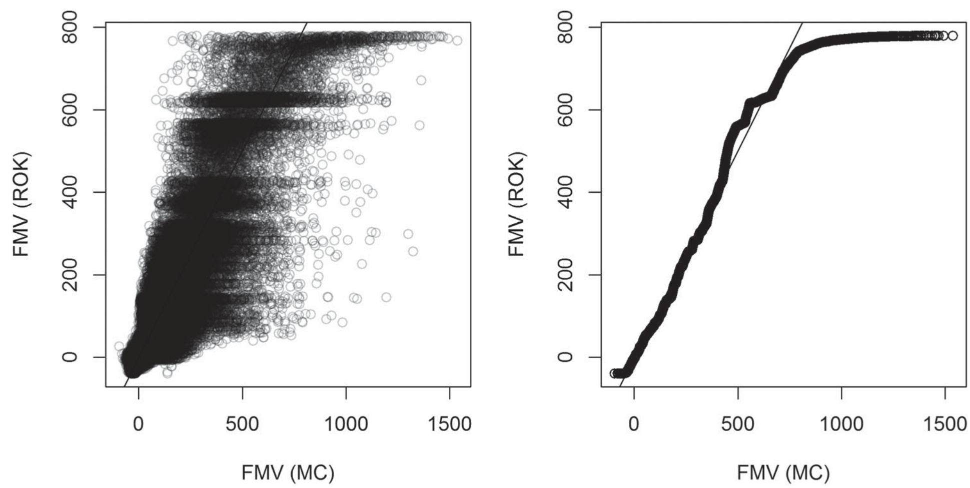
In this appendix, we provide some test results about the sensitivity of the experimental design methods on the three metamodels. Gan and Valdez (2016) compared five experimental design methods for the GB2 regression model and found that the data clustering method and the conditional Latin hypercube sampling method produce the most accurate results. Here we compare two additional experimental design methods—conditional Latin hypercube sampling and random sampling—for the three metamodels.

Table B.1 shows the performance of the three metamodels based on the two experimental design methods. The results are mixed. Table B.1(a) shows that GB2 regression works the best when conditional Latin hypercube sampling was used to select representative policies. Table B.1(b) shows that rank order kriging produced the best result in terms of *CCCQ* when the random sample was used. In terms of *PE* and *ME*, rank order kriging did not perform as well as the other two metamodels when conditional Latin hypercube sampling and random sampling were used to select representative policies. In terms of R^2 and *CCCQ*, rank order kriging outperformed ordinary kriging.

Figure B.1 shows the scatter plot and the Q–Q plot produced by rank order kriging with conditional Latin hypercube sampling. These plots indicate that rank order kriging does not work well with conditional Latin hypercube sampling. Figure B.2 shows the scatterplot and the QQ plot produced by rank order kriging with random sampling. Comparing

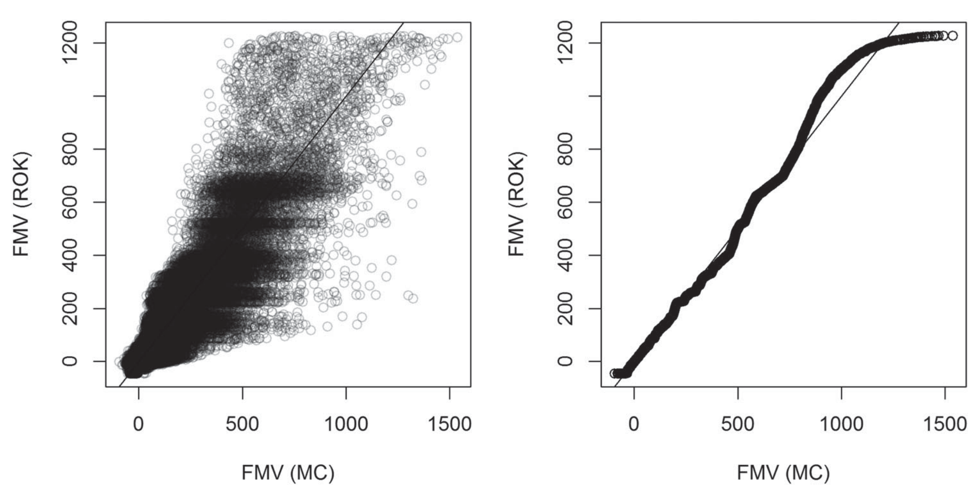
TABLE B.1
Accuracy of the Metamodels When $k = 340$ Representative VA Policies Were Selected by the Two Different Experimental Design Methods

(a) Conditional Latin hypercube sampling				
	PE	*ME*	R^2	*CCCQ*
Rank order kriging	–0.0720	–4.8676	0.8156	0.9791
Ordinary kriging	–0.0365	–2.4654	0.7965	0.9256
GB2 regression	0.0331	2.2375	0.8238	0.9971
(b) Random sampling	*PE*	*ME*	R^2	*CCCQ*
Rank order kriging	–0.0335	–2.2673	0.8396	0.9954
Ordinary kriging	0.0117	0.7876	0.8099	0.9520
GB2 regression	0.0172	1.1600	0.8729	0.9887



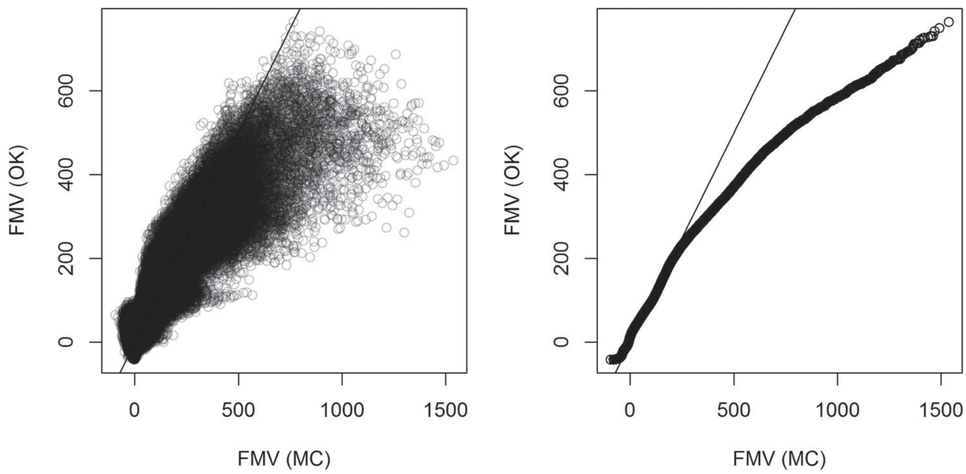
(a) (b)

FIGURE B.1. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by Rank Order Kriging When $k = 340$ Representative VA Policies Were Selected by Conditional Latin Hypercube Sampling.



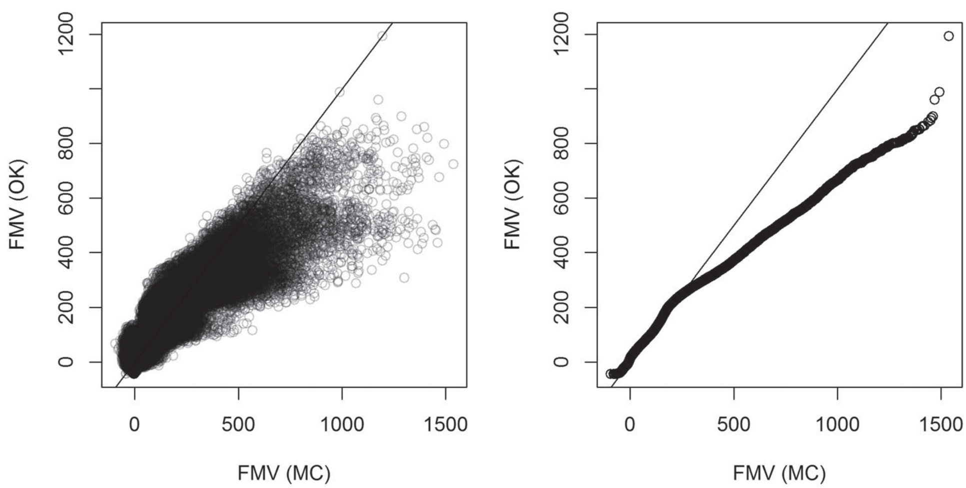
(a) (b)

FIGURE B.2. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by Rank Order Kriging When $k = 340$ Representative VA Policies Were Selected by Random Sampling.



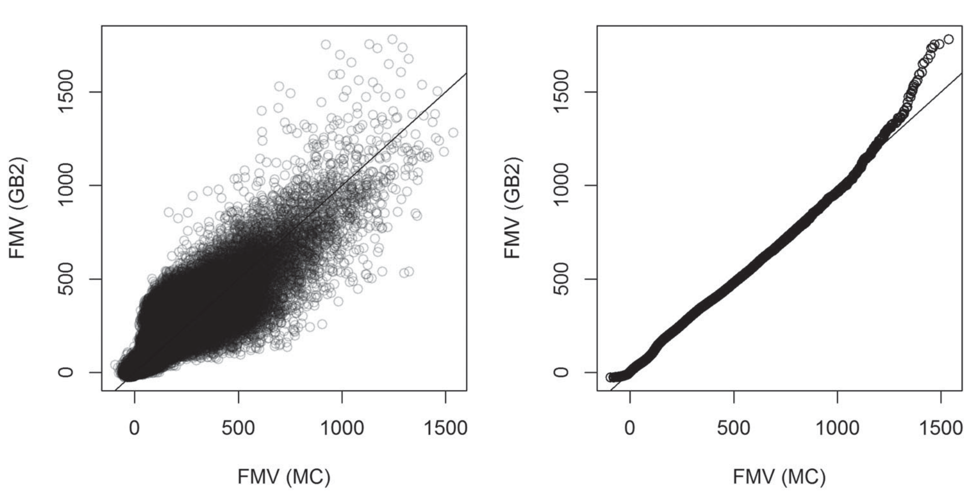
(a) (b)

FIGURE B.3. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by Ordinary Kriging When $k = 340$ Representative VA Policies Were Selected by Conditional Latin Hypercube Sampling.



(a) (b)

FIGURE B.4. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by Ordinary Kriging When $k = 340$ Representative VA Policies Were Selected by Random Sampling.



(a) (b)

FIGURE B.5. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by GB2 Regression When $k = 340$ Representative VA Policies Were Selected by Conditional Latin Hypercube Sampling.

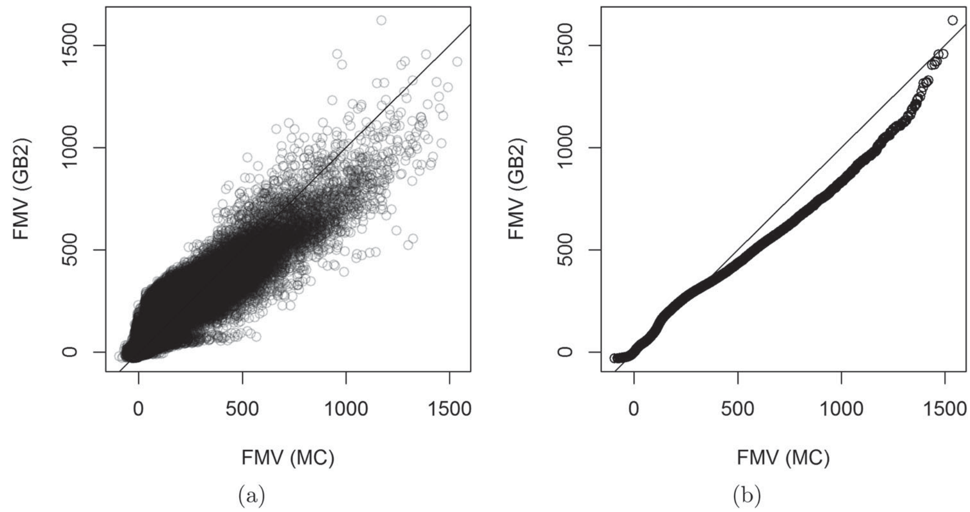


FIGURE B.6. Scatter and Q–Q Plots of the Fair Market Values Calculated by Monte Carlo and Those Estimated by GB2 Regression When $k = 340$ Representative VA Policies Were Selected by Random Sampling.

Figure B.2 with Figure B.1, we see that rank order kriging produced better results when random sampling was used to select representative policies.

Figures B.3, B.4, B.5, and B.6 show the scatterplots and the Q–Q plots produced by ordinary kriging and GB2 regression with the two experimental design methods. These figures show that ordinary kriging does not fit the right tail well and GB2 regression produced better results in both cases.