

Park, Jaewon; Shin, Minsoo

Article

An approach for variable selection and prediction model for estimating the Risk-Based Capital (RBC) based on machine learning algorithms

Risks

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Park, Jaewon; Shin, Minsoo (2022) : An approach for variable selection and prediction model for estimating the Risk-Based Capital (RBC) based on machine learning algorithms, Risks, ISSN 2227-9091, MDPI, Basel, Vol. 10, Iss. 1, pp. 1-20, <https://doi.org/10.3390/risks10010013>

This Version is available at:

<https://hdl.handle.net/10419/258324>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.


If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Article

An Approach for Variable Selection and Prediction Model for Estimating the Risk-Based Capital (RBC) Based on Machine Learning Algorithms

Jaewon Park  and Minsoo Shin *

Department of Management Information System, School of Business, Hanyang University, Seoul 04763, Korea; jwhanyang@hanyang.ac.kr

* Correspondence: minsooshin@hanyang.ac.kr; Tel.: +82-2-2220-1988

Abstract: The risk-based capital (RBC) ratio, an insurance company's financial soundness system, evaluates the capital adequacy needed to withstand unexpected losses. Therefore, continuous institutional improvement has been made to monitor the financial solvency of companies and protect consumers' rights, and improvement of solvency systems has been researched. The primary purpose of this study is to find a set of important predictors to estimate the RBC ratio of life insurance companies in a large number of variables (1891), which includes crucial finance and management indices collected from all Korean insurers quarterly under regulation for transparent management information. This study employs a combination of Machine learning techniques: Random Forest algorithms and the Bayesian Regulatory Neural Network (BRNN). The combination of Random Forest algorithms and BRNN predicts the next period's RBC ratio better than the conventional statistical method, which uses ordinary least-squares regression (OLS). As a result of the findings from Machine learning techniques, a set of important predictors is found within three categories: liabilities and expenses, other financial predictors, and predictors from business performance. The dataset of 23 companies with 1891 variables was used in this study from March 2008 to December 2018 with quarterly updates for each year.

Keywords: life insurance companies; Bayesian Regulatory Neural Network; Random Forest algorithms; RBC ratio; corporate sustainable management; Machine learning



Citation: Park, Jaewon, and Minsoo Shin. 2022. An Approach for Variable Selection and Prediction Model for Estimating the Risk-Based Capital (RBC) Based on Machine Learning Algorithms. *Risks* 10: 13. <https://doi.org/10.3390/risks10010013>

Academic Editor: Mogens Steffensen

Received: 19 November 2021

Accepted: 29 December 2021

Published: 4 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

Life insurance is one of the most important financial products for sustainable finance and sustainability in financial consumers. Life insurance companies deal with financial products (i.e., life insurance policies) that are critically related to a household's financial security after a cataclysmic event (i.e., loss of a breadwinner) (Rejda 2008; Thoyts 2010). Any failure in supervising the life insurance companies can be critically associated with households' financial well-being (Financial Supervisory Service 2017). Specifically, as a financial product for survivors, life insurance serves the primary function of securing a certain amount of wealth, which can help recover a financial loss incurred by the premature death of a breadwinner (Rejda 2008; Thoyts 2010). Therefore, insurance policyholders, financial educators, policymakers, and insurers need to understand which factors are associated with life insurance companies' stability and reliability to build a sustainable financial environment.

Korean life insurance companies' current stability and reliability can be assessed using the risk-based capital (RBC) ratio method. Essentially, the RBC can be used as a screening tool by investors and consumers and as a policy tool for regulation. Insurance business law and the Insurance Business Supervision Regulation enforce that the RBC ratio should

be 100% or higher as a measure of the equity ratio ([Enforcement Decree of the Insurance Business Act § 65-2 2019](#); [Insurance Business Supervision Regulation § 7-17 2020](#)).

The RBC ratio is measured by comparing two components: how much capital a life insurance company has (i.e., available capital, AC) and how much capital a life insurance company should have (i.e., required capital, RC). AC can be measured using a simple calculation of the total capital that a life insurance company currently possesses (i.e., core capital and supplementary capital). However, RC should be measured considering the potential risks that a life insurance company should prepare for (i.e., insurance risk, interest rate risk, credit risk, market risk, and operational risk). As will be shown in the examination of its theoretical background, the RBC ratio is calculated using a specific mathematical calculation. Based on this calculation, some direct factors (i.e., core capital, supplementary capital, interest rate risk, and market risk) are obviously expected to show a life insurance company's RBC ratio for the next period. However, in terms of risk and risk management, the calculations of total risk for RC are not quite predictable or expectable, due to the market and the available capital's volatility in the life insurance field ([Heo 2020](#)). Therefore, it is important to discover a prediction model for financial failure by capturing determinants through strong predictors that significantly affect the financial soundness of life insurance companies.

1.2. Research Gap

In conventional studies about the financial market, conventional approaches such as linear estimation were generally utilized to predict the next period's outputs (e.g., RBC ratio). The conventional approach to predicting an output utilizes significant independent variables' marginal effects. Statistically, the conventional approach is limited because it utilizes the coefficients of predictors that are statistically based on unique variance among output variables and predictors ([Chatfield and Xing 2019](#)). In other words, the conventional approach disregards the covariances among the predictor variables. This statistical disregard of covariance makes the conventional approach strong in terms of explanations but weak in terms of predictions ([Heo 2020](#)). The coefficient found using conventional approaches strongly explains how predictors are associated with outputs. However, the patterns to predict the outputs are limited by the coefficients of determination (i.e., R^2 , adjusted R^2 , pseudo R^2 , and similar alternative indicators). As a result, the conventional approach is strong in terms of finding the explanatory power of a specific predictor but weak at finding a set of efficient predictors ([Heo 2020](#)). On the other hand, Machine learning from the computer science field is robust in terms of finding patterns and predicting outputs ([Ye 2014](#)). Due to the fact that Machine learning techniques use a total sum of predicting weights from predictors such as using neural network algorithms, the predictions undertaken using Machine learning techniques were known to be efficient and strong in terms of prediction and classification ([Kudyba 2014](#); [Linoff and Berry 2011](#); [Thompson 2014](#); [Ye 2014](#)). Machine learning's strengths in prediction can be adapted to finance or related research fields (e.g., [Bosarge 1993](#); [Heo 2020](#); [Heo et al. 2020](#)). As noted above, Machine learning techniques are weak in terms of explaining how each predictor changes the output. This is the reason why Machine learning techniques are known as black-box modeling; the modeling is not free from criticism regarding its explanatory function ([Rudin 2019](#)). However, considering that this paper's research purpose is to find a set of predictors to increase prediction rate, Machine learning techniques are acceptable and sometimes more helpful than conventional statistical approaches. One of the most important steps of Machine learning (ML) techniques is the feature selection to determine a set of significant inputs in the large number of variables ([Muttill and Chau 2007](#)). Selecting an appropriate set of significant inputs for a data-driven model provides obvious advantages. Regarding these points, this study is based on a data-driven dynamical systems approach with Machine learning, in which a mathematical framework is provided to describe the modeling of the rich interactions for data analysis and prediction between quantities that co-evolve over time ([Brunton and Kutz 2019](#)). Recently, data-driven models, including

Machine-learning (ML) techniques, have been employed across a diverse range of complex systems such as finance, brains, robotics, and autonomy. Heo (2020) suggests that all possible data is considered for the prediction of performance in a dynamic ecological systemic framework, because all the factors from multiple systems have a mutual influence on the outcomes. Hence, ML techniques, such as Artificial neural networks (ANNs), are applied as dynamic ecological factors to data-driven approaches for a large number of predictors (Heo 2020).

Machine learning spreads rapidly to various areas, and Machine learning techniques are applied to many areas when used as a prediction tool. Researchers and practitioners use Machine learning to learn data patterns and apply them to various situations. Machine learning research in the modern financial market was utilized to improve the ability to assess risk exposure and detect early warning indicators using a computer science approach (Kou et al. 2019). Beutel et al. (2019) concluded that Machine learning helped to predict the banking crisis using a comprehensive dataset covering the system's financial crisis for 15 advanced economies over the past 45 years. In the empirical literature, the prediction of bank failure focuses on identifying leading indicators that contribute to reliable early warning systems. In addition, accounting-related literature applying Machine learning tools to predict the quality of accounting numbers is increasing (Liu et al. 2021). Barboza et al. (2017) compared traditional statistical methods and Machine learning using the financial data of North American firms. According to Barboza et al., the performance of Machine learning models (i.e., support vector Machines, bagging, boosting, and Random Forest) was compared the performance of discriminant analysis, logistic regression, and Neural networks. The result showed that Machine learning models had higher levels of prediction accuracy than the traditional models. Primarily, Random Forest provided better accuracy and error rates than the other models.

The literature also supports the utility of using Machine learning in finance, as it improved empirical predictions. For instance, Kartasheva and Traskin (2013) used the Random Forest (RF) classification algorithm for insurers' insolvency predictions of the U.S. property-casualty insurance companies, using the characteristics and information about the insolvencies of insurance. Kartasheva and Traskin compared logistic regressions, which are commonly utilized for the predictions of insurers' financial conditions. The results showed that the RF methodology has a higher prediction quality than the existing method and ranked variables' importance to provide a practical guide for monitoring regulators and market participants. One study (Hutagaol and Mauritsius 2020) used Machine learning to access risk level predictions of potential customers of the largest life insurance company in Indonesia. In order to select the best model, three methodologies, such as the Support Vector Machine (SVM) algorithm, naïve Bayes algorithm, and Random Forest (RF) were evaluated according to accuracy, recall (sensitivity), and precision. RF showed the highest precision, accuracy, and sensitivity. Similarly, Ding et al. (2020) compared the model-generated estimates of managers' predictions of financial reports to evaluate whether the Machine learning model outperformed managers in terms of predictive accuracy. Ding et al. used the annual report of US-based property and casualty insurance companies filed with the National Association of Insurance Commissioners (NAIC) covering the period from 1996 to 2017 and compared four popular Machine learning algorithms (Linear Regression, Random Forest, Gradient Boosting Machine, and Artificial neural networks) to predict insurance losses. The results suggest that Machine learning techniques can be beneficial to managers and auditors in improving accounting estimates. Additionally, the study found that the Random Forest algorithm showed the best accuracy and prediction among those examined. The 15 most powerful and influential variables identified by the Random Forest algorithm play an essential role in generating model predictions.

Further literature found that Machine learning was useful in finance to generate improved predictions using various Machine learning methods. For instance, Serrano-Cinca and Gutiérrez-Nieto (2013) analyzed the 2008 financial statements from 8293 banks using a mathematical estimation method to predict the 2008 U.S. bank crisis. The paper identified

significant issues of bank financial soundness based on Partial Least Square Discriminant Analysis (PLS-DA), which was compared with eight algorithms which were previously widely used in bankruptcy prediction studies (i.e., Linear Discriminant Analysis (LDA), Logical Regression (LR), Multilayer Perceptron (MLP)), K-Nearest Neighbors (KNN), Naive Bayes (NB), and Support Vector Machine (SVM)). The PLS-DA results were very similar to those obtained by Linear Discriminant Analysis and Support Vector Machine and showed that this technique was helpful for dimension reduction and data selection. [Petropoulos et al. \(2020\)](#) used a series of modeling techniques to predict bank insolvency from a sample of US-based financial institutions with datasets of more than 175,000 records, which provided quarterly information for seven years from 2008 to 2014 in the Federal Deposit Insurance Corporation (FDIC) database. The empirical results showed that Random Forest has superior out-of-sample and out-of-time predictive performance, and Neural networks showed almost the same performance as RF in out-of-time samples. This result was reached following a comparison with broadly used bank failure models and advanced Machine learning techniques such as Logistic Regression (LogR), Linear Discriminant Analysis (LDA), Random Forest (RF), Support Vector Machines (SVMs), Neural Networks (NNs), and Random Forest of Conditional Inference Trees (CRF).

In terms of RBC prediction, Machine learning methods were used in previous literature to find the usefulness of Artificial neural networks ([Hsiao and Whang 2009](#)). [Hsiao and Whang \(2009\)](#) conducted a study of 15 domestic companies and 10 overseas life insurers with complete financial data to evaluate the financial situation of selected samples in Taiwan and monitor their ability to pay using the RBC ratio and the total financial index (TPI) of the CAMEL-S model. To this end, research using data sources concerning life insurance companies' annual financial reports, previous related studies, and archive information collected from the Taiwan Rating Corporation (TRC) database found that it was useful to use Artificial neural networks to predict RBC ratios compared with conventional discriminant methods. [Liu et al. \(2021\)](#) investigated Machine learning and statistical techniques used in the literature in more than 60 major papers about insolvency prediction. The study found that use of Machine learning-based models is increasing rapidly, and most Machine learning approaches, such as Random Forest, have compared to traditional approaches for their significant predictive ability. However, more research into Machine learning study is required, considering various leading indicators under regulatory differences between countries. For example, based on the literature, almost half of the Machine learning papers used U.S. bank data and the other half the data of several European countries, but few studies on Asia Pacific countries exist. Furthermore, research about other financial institutions, such as insurance companies, is relatively limited, while research on bankruptcy in the banking sector has developed well.

Overall, these studies greatly complement our research, as there are potential factors to be associated with RBC as external factors. Specifically, by employing Machine learning algorithms, the potential factors can be found from a large dataset and be used to improve RBC prediction. As results, unlike many studies that have focused on comparing the predictive power of several Machine learning approaches, the purpose of our study is to find a set of important predictors to estimate the RBC ratio of all Korean life insurance companies that have not yet been studied. For this study, The Random Forest algorithm, and the Bayesian Regulatory Neural Network (BRNN) were utilized for variable selection and prediction. The Random Forest algorithm was adopted in many studies considering variable selection and prediction power and handling large variables ([Genuer et al. 2010](#)). The Bayesian Regulatory Neural Network (BRNN) is one of the neural network algorithms used to predict outputs with some strength ([Sariev and Germano 2019](#)). For instance, models are not overly trained or not overly fitted and are effective types of Neural networks ([Burden and Winker 2008](#)). Specifically, the overtraining issue was adjusted because the BRNN stopped performing the estimating procedure when the neural network reached the least parameters ([Lau et al. 2009](#)). Therefore, the BRNN is an effective tool to compare

against the conventional statistical model ability to predict outcomes, because BRNN does not exaggerate the prediction power of Machine learning.

1.3. Research Purpose

In terms of the subject of this study (i.e., the RBC ratio), the RBC system has a basic structure with a formula for standard calculations of the financial soundness of each risk factors; it is not made up solely by summing the individual risks but calculated using all types of risk coefficients to effectively consider the total risks of an insurer ([Financial Supervisory Service 2017](#)). The International Association of Insurance Supervisors recommends reflecting the variance effect of individual risks rather than using a simple summation method when calculating the total required capital in insurance core principles and a common framework for the supervision of internationally active insurance groups ([International Association of Insurance Supervisor 2021](#)). This means that many components used to calculate the RBC are affected by hidden external effects, although the RBC seems to be calculated using an exact variable. It is noteworthy that the study utilized the Machine learning method. As discussed, various external factors may be associated with the RBC ratio, and the exact variables were possibly middle stations used to calculate the RBC ratio as the final output. Based on this assumption, finding a variable selection using a data-driven dynamical systems approach with Machine learning that looks at patterns and predicts as many variables as possible in order to find hidden external factors that influence the RBC is an important process ([Beutel et al. 2019](#)). The purpose of this study is to find the variable set that will make better predictions of the RBC ratio. Based on the selected variables, other researchers may investigate the external factors of the process of calculating the RBC ratio. Therefore, this study delimits the purpose for variables selection, instead of being a comparison of various Machine Learning techniques. Generally, previous literature with Machine Learning techniques compared various versions of Machine Learning and found a specific algorithm's usefulness in predicting an output. However, this study has a specific purpose to find the variable list to be associated with RBC ratio prediction so that other researchers may see them as the possible research topics. To achieve this purpose, variable selection algorithms were used to find the potential external variables, and a neural network algorithm was used to confirm the prediction accuracy as a cross-validation method.

As a result, this study set out to find a set of important predictors to estimate the RBC ratio of life insurance companies across a large number of variables (1891), including key finance and management indices collected from all Korean insurers quarterly under regulation for transparent management information. We employ two extensively used data-driven models for selecting significant input variables, considering hidden external factors. Initially Random Forest algorithms and then Bayesian Regulatory Neural Network (BRNN) were used to predict the performance of the selected variables. The results in this paper highlight the most important factor for an indicator of sustainability management of insurance companies. The findings will contribute not only to the safety and soundness of individual insurance companies, but also the stability of the financial system by monitoring selected significant variables in advance. It is also expected to help encourage the building of a sustainable financial system for long-term value creation with social responsibility for consumer safety.

The significance of this study can be summarized as follows. First, the results can identify the order of the most important variables in RBC prediction for the sustainability of Korean life insurance companies. Second, it can be used as a reference for the monitoring and forecasting of risk management so that sustainable insurance for corporate social responsibility can be maintained. Third, the dataset of a bigger dimension of Korean life insurance companies, which has not yet been predicted, was used for a prediction model. The remainder of the paper is divided as follows: An overview of the relevant literature is provided in Section 2, with a detailed description of the history of the RBC ratio of insurance companies in South Korea. The theoretical Background of the RBC is covered in Section 3, followed by the statistical background of Machine Learning algorithms in

Section 4. The research methodology is outlined in Section 5, with a presentation of the results in Section 6. Finally, the conclusion and discussion are detailed in Section 7.

2. The History of the RBC Ratio of Insurance Companies in South Korea

During the global financial crisis in 2008, large insurance companies, such as American Insurance Group Inc. (AIG), collapsed (CBS 2008), and the crisis affected the financial systems of many countries across the world. Therefore, in many countries, financial supervisory governmental entities (e.g., the Financial Supervisory Service in South Korea) recognized that the existing system alone was insufficient to prepare for the next potential risk (National Association of Insurance Commissioners n.d.). Based on this need for a new regulatory system for insurance companies, financial supervisory experts tried to improve the system of determining the financial soundness of insurance companies. In the case of the regulation of equity capital for Korean life insurance companies, a simple allowance system has been created, in which a certain ratio was applied to the reserve for past liabilities (Financial Supervisory Service 2017). In line with the emerging global insurance regulations and supervisory approach, the Financial Supervisory Service decided to adopt RBC regulations in April 2009. There was a two-year grace period to ensure that insurance companies had adequate preparation time to incorporate any changes. The new regulatory evaluation method, the RBC, came into effect in April 2011.

The RBC ratio method aims to ensure that financial institutions have enough capital to deal with unexpected losses (National Association of Insurance Commissioners 2020). The RBC ratio method is an evaluation method for insurance companies that measures the amount of risk inherent in the corresponding capital held by an insurance company. It assesses capital adequacy through a calculation using two components: the available capital and the required capital. In this assessment, the available capital is a risk buffer that allows the insurance company to maintain its solvency margin¹ so that it can handle unexpected losses. The required capital is the necessary capital based on the insurer's exposure to market risks, such as interest rate risk, credit risk, and operational risk. The RBC ratio of available capital divided by required capital is used as a basis for rapid corrective action and as a management or risk assessment indicator for Korean life insurance companies (Financial Supervisory Service 2017).

Regulations and Maintenance of Financial Soundness

The current RBC ratio method used to regulate insurance companies in Korea was activated as a proactive Act with a practical policy for ensuring insurance companies' paying ability as well as companies' sound management (Enforcement Decree of the Insurance Business Act § 65 2019, Insurance Business Act § 123 2019). According to the proactive law and policy, insurance companies must align their business operations to comply with the standards for financial soundness, such as adequacy of capital, soundness of assets, and other necessary requirements. For instance, insurance business law states that the RBC ratio should be 100% or higher as a measure of the equity ratio (Enforcement Decree of the Insurance Business Act § 65-2 2019). As of the end of each final quarter in a year, insurance companies must keep their solvency margin ratios² over 150% (Insurance Business Supervision Regulation § 2-6-3 2020). In addition, as a screening criteria for merger authorization, insurance companies should keep the ratio of allowances to be paid at 100% or higher after any merger (Insurance Business Supervision Regulation § 7-36 2020). Furthermore, the regulation states that the actual condition evaluation system for risk assessment (RAAS: Risk Assessment and Application System) should be the same as the risk standard for the actual management of financial conditions for supervising the insurance companies' soundness (Insurance Business Supervision Regulation § 7-14 2020). The capital surplus ratio is also included in the measurement items of the capital adequacy evaluation section of the overall evaluation.

To ensure the insurance companies follow the threshold level of the RBC ratio, timely corrective measures are adjusted in the regulations. The timely corrective measures can

be broadly divided into two: requirements for equity ratio and management condition evaluation. If the allowance ratio is more than 50% and less than 100%, management improvement recommendations should be made to implement necessary measures ([Insurance Business Supervision Regulation § 7-17 2020](#)). If the allowance ratio is more than 0% and less than 50%, requests for management improvement should be made ([Insurance Business Supervision Regulation § 7-18 2020](#)). When the allowance ratio is less than 0%, the commission shall order the insurance company to take necessary measures to implement orders for management improvement ([Insurance Business Supervision Regulation § 7-19 2020](#)).

To sum up, the Korean government and the related supervisory entity (i.e., Financial Supervisory Service) utilized the RBC ratio as the main function to secure the stability of the insurance industry, including life insurance companies, which is explained conceptually below in Section 3, which covers the theoretical background of the RBC ratio.

3. The Theoretical Background of the RBC Ratio

As shown below in Figure 1, the RBC ratio has two main components. The first component, available capital (AC), is calculated from the sum of core capital and supplementary capital after itemized deduction. The second component, required capital (RC), is the total risk of an insurance company. Overall, the ratio between the AC and the RC of a company are calculated as the RBC ratio, and the timely corrective measures and risk assessment and application system (RAAS) are used as an adjustment component for the RBC ratio ([Insurance Business Supervision Regulation § 7-1 2020](#)). The detail of RBC ratio calculation is shown on Table A1 in Appendix A.

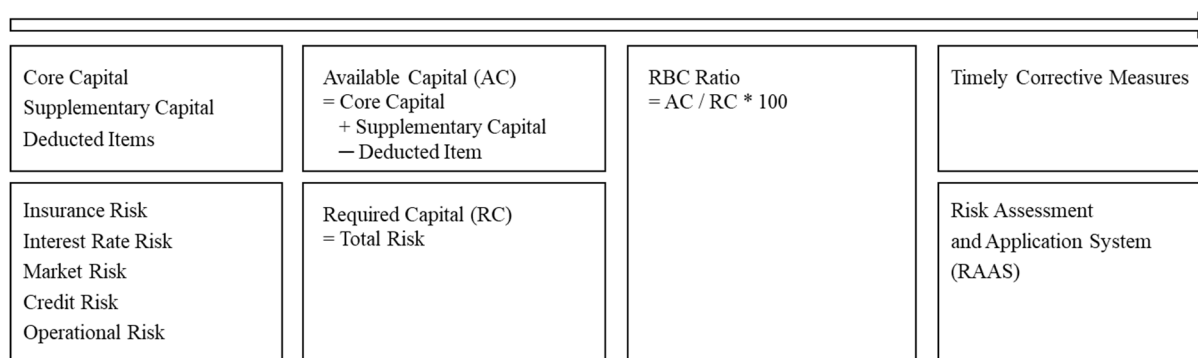


Figure 1. Calculating RBC ratio. Source: [Financial Supervisory Service \(2017\)](#).

In Figure 1, risks are categorized as insurance risks, interest rate risks, credit risks, market risks, and operational risks ([Financial Supervisory Service 2017](#)). Insurance risk denotes the risk of loss due to an unexpected increase in loss ratio, which is determined from the loss ratio and claims reserve level. Interest risk is the risk of negatively affecting the financial condition due to a decrease in the value or net assets, which can be caused by interest rate fluctuation. Credit risk is the risk of loss due to the default of the debtor or the default of the counterparty. Market risk is the risk of loss due to changes in asset value caused by changes in marketplaces, such as stock price, interest rate, and exchange rate. Operational risk is the risk of loss due to improper internal procedures, personnel, systems, or external events, which is determined by the adequacy of internal control and preventive measures against incidents.

The required capital (RC) is calculated by summing up all types of risks to effectively consider the total risks of an insurer. However, in the RBC calculation methods, required capital is not simply the sum of the individual risks. Instead, it is calculated from the total risks by reflecting the dispersion effect ([Financial Supervisory Service 2017](#)). To make the RBC ratio more realistic, the domestic RBC system refines the correlation coefficient (See Appendix A Table A2) of individual risk amounts to calculate the integrated risk corresponding to the risk characteristics of each insurance company. As a result, the

RBC ratio is calculated using the following Function (1) ([Insurance Business Supervision Regulation § 7-2 2020](#)):

$$RBC\ ratio = AC/RC \quad (1)$$

4. Statistical Background: Machine Learning Algorithms

4.1. Random Forest

The family of Random Forest algorithms is well known as being a useful feature for prediction and classification, which are specifically useful when there is a significant number of predictors ([Genuer et al. 2010](#)). Random Forest algorithms are known as an ensemble method that combines the decision tree process into Machine learning techniques ([Breiman 2001](#)). In Random Forest algorithms, multiple partitions of decision trees are implemented with bootstrap sampling so that the algorithms can produce a list of important variables for improved prediction, following Function (2) and Function (3) ([Biau 2012](#)):

$$r_n(X, \Theta) = \left\{ \sum_{i=1}^n Y_i 1_{[X_i \in A_n(X, \Theta)]} / \sum_{i=1}^n 1_{[X_i \in A_n(X, \Theta)]} \right\} 1_{E_n(X, \Theta)} \quad (2)$$

$$r_n - \text{bar} = E_{\Theta}[r_n(X, \Theta)] \quad (3)$$

where, $r_n(X, \Theta)$ is a randomized tree; X denotes the set of predictors; Θ is the independent predictor of X ; Y_i denotes all outputs predicted by random partitions of X (i.e., X_i); X_i is the random partition of X so that $A_n(X, \Theta)$ means the partial data frame of X ; $E_n(X, \Theta)$ is the entity of partial matrix which is not equal to zero (i.e., empty information will make the function as zero); $r_n\text{-bar}$ means the estimated Random Forest; and E_{Θ} is a matrix of variables.

4.2. Random Forest Boruta

Based on the basic model of Random Forest, the algorithms in the Random Forest Boruta packages consist of this sequential order ([Kursa and Rudnicki 2010](#)): (a) make a random partition in the data and duplicate the random partitions multiple times; (b) execute the Random Forest classifier to compute maximized z-scores across partitions; (c) find the maximized z-scores to reach the threshold to select important predictors; (d) compare each predictor's z-score with the maximized z-score; and (e) repeat steps (a) through (d) until all predictors are categorized as important predictors or unimportant predictors. By completing this procedure, the Boruta algorithms may classify large numbers of inputs (i.e., predictors) into two categories: important predictors and unimportant predictors.

4.3. Random Forest Recursive Feature Elimination

When the number of predictors is comparatively large as in this study (i.e., 1890 predictors at the initial stage), a combination of the same family of algorithms (e.g., Random Forest) can be empirically utilized to find the best rank of predictors ([Zhou et al. 2014](#)). Specifically, Random Forest RFE is an algorithm used to create a rank of predictors in an ascending order with predicting weights ([Guyon et al. 2004](#)) using the following procedure: (a) the dataset is partitioned to partial matrix at the first procedure; (b) using the partitioned dataset, the importance of predictors to output is computed; and (c) by repeating (a) and (b), the least important predictors are sorted out and the important predictors ordered. Finally, the Random Forest RFE shows a list of the predictors in ascending order with predicting weights.

4.4. Bayesian Regulatory Neural Network

BRNN is a feed-forward type of Artificial Neural Network (ANN), which means that data used in the initial process was only used for estimating the output going forward ([Sariev and Germano 2019](#)). As a result of the one-directional process, BRNN estimates efficient outputs from nonlinear data without any overfitting or overtraining ([Burden and](#)

Winker 2008). As the baseline of BRNN, the function for ANN with multiple hidden layers is similar to the below Function (4) (Heo 2020):

$$y = f[\sum_{j=1}^m \omega_j * (\sum_{i=1}^n \omega_{ij} X_i)] \quad (4)$$

where y denotes the predicted output; X_i are input variables (i.e., predictors); and ω is the weights for each variable. Using Function (4) for ANN, the Bayesian rule regularized the estimating procedure to make the least parameters, as in below Function (5) (Lau et al. 2009):

$$F(y, W) = \alpha E_s + \beta E_W \quad (5)$$

where E_s denotes the neural network's prediction error by calculating the mean of the sum of squares of the network's errors; and E_W is the network's prediction of weights by calculating the mean of the sum of squares of the network's weights (ω). Therefore, α denotes the performance function parameters for the Neural networks' errors, and β represents the performance function parameters for the Neural networks' weights. When β is larger than α , the overfitting issue has been solved. Therefore, the Neural networks operate until the parameter the of network error is smaller than the parameter of the network weights. As the functions indicate, the BRNN estimates the output by minimizing the overfitting issue. Therefore, there are few possibilities of exaggerating the performance of Machine learning. This is the reason why, in this study, the BRNN has been utilized as the method to confirm the performance of the variables found by Random Forest RFE.

4.5. Ordinary Least Squared Modeling

Ordinary least squares (OLS) is conventional linear estimation when compared to BRNN, the Machine learning method, by RMSE and MAE to evaluate which model has better prediction for the RBC ratio in the next quarter. The function of OLS is similar to Function (6) below.

$$y = a + \sum_{i=1}^n (b_i X_i) + e \quad (6)$$

where y is the dependent variable; a is the constant of a model; b_i are regression coefficients; and X_i denotes 29 predictors selected as explanatory variables of the model.

5. Method

5.1. Data

Data were pooled from the Financial Statistics Information System (FISIS), which was collected by the Financial Supervisory Service (FSS) in South Korea. This is public data with one source merged from the FISIS (Financial Statistics Information System) by Korean Financial Supervisory Service. This is based on the reports created under the responsibility of the financial company within three months of the end of each quarter. (Data source link; Retrieved 10 December 2020, from <http://efisis.fss.or.kr/fss/fsiview/indexw.html>). FSS is a quasi-governmental organization that supervises and oversees all types of financial institutes (i.e., banks, insurers, investment companies, non-bank financial institutes, and other financial institutes) in South Korea (Financial Supervisory Service n.d.). As a tool for supervising all types of financial institutes in South Korea, FSS collects various information about financial institutes, including general information (e.g., number of officers, employees, number of branches, etc.), financial information (e.g., balance statements, profit and loss accounts, real estate assets, loan information, income, and expenses statements, etc.), key management indices (e.g., capital adequacy, profitability, liquidity, etc.), and major business activities (e.g., premium revenue, claims paid, refunds, net operating expenses, etc.). In addition, these data are publicly available to all taxpayers in South Korea, because the FSS is operated by taxpayers. As a result, the dataset is valid and reliable, because the data is collected for the purpose of supervising all types of financial institutes.

From the FISIS, there were 24 companies (15 domestic life insurance companies and 9 foreign life insurance companies) registered to FSS. However, one of the domestic life

insurance companies, Kyobo Life Planet Insurance Company, was founded in 2013, so the important data (i.e., RBC) for that company was not obtained in 2013. Therefore, the datasets of 23 companies were used in this study from March 2008 to December 2018 with quarterly updates.

5.2. Output Variable: The RBC Ratio of the Next Quarter

The various information about financial institutes included 1891 variables, which are general information, financial information, key management indices, and major business activities. The output variable to be predicted by the Machine learning algorithms was the RBC ratio of the next quarter, which, as explained above, is an important index for measuring the soundness of life insurance companies. Except for the output variable, the other 1890 variables were (a) used, (b) sorted, and (c) ranked by the above analytic algorithms to predict the RBC ratio of the next quarter. In the dataset, all 1891 variables were collected on a quarterly basis. As shown in below Function (6), functions of input variables were utilized to predict the RBC ratio in the next quarter:

$$f_j[\Sigma(Xit)] \rightarrow Y_{t+1} \quad (7)$$

where, f_j denotes four algorithms' functions (i.e., Random Forest Boruta, Random Forest RFE, BRNN, and OLS) that are explained below in Table 1; Xit are the input variables in the quarter of t ; and Y_{t+1} is the output variable in the next quarter $t+1$.

Table 1. Analytic Algorithms in the Study.

	Analytic Algorithms	Note
Stage 1: Feature Selection	Random Forest: Boruta RFE	Unimportant variables are sorted out to predict the next quarter's RBC. The optimal number of variables to predict the next quarter's RBC was found.
Stage 2: Predicting Stage	Validating by ML: BRNN Comparing method: OLS	The next quarter's RBC with the optimal number of variables was predicted using BRNN. The next quarter's RBC with the optimal number of variables was predicted using OLS.

Note. RFE is Recursive Feature Elimination; RBC denotes Risk-Based Capital; BRNN is the Bayesian Regulated Neural Network; OLS is Ordinary Least Squared modeling; ML is Machine Learning. Source: Authors' construction.

5.3. Analytic Algorithms

The purpose of this study was to find key variable for RBC forecast among 1890 variables that make up the major financial and non-financial indicators of life insurance companies.

In the Machine learning techniques, it is known that the Machine learning algorithm shows lower prediction accuracy when using large numbers of variables compared with the optimal number (Kohavi and John 1997). Thus, the feature selection with a set of significant inputs is one of the most important steps when evaluating a large number of variables (Muttill and Chau 2007). In this paper, Machine learning algorithms are utilized following three steps according to a practical point of view: First, Random Forest Boruta was used to classify the number of variables into two categories: important predictors and unimportant predictors. Second, Random Forest RFE was used to rank the variables in ascending order with predicting weights. Third, the Bayesian Regularized Neural Network (BRNN) was utilized to predict the performance of a set of significant input variables selected by Random Forest algorithms.

As shown in Table 1 and Figure 2 below, the analytic algorithms followed two stages: (a) a feature (i.e., variable) selection stage and (b) the stage for predicting the RBC ratio in the next quarter. The first stage used the algorithms to select the important variables from the original 1890 variables (i.e., various information from 24 companies). The first stage consisted of two sub-stages using Random Forest algorithms, Random Forest Boruta, and Random Forest RFE. Random Forest Boruta used a function to sort out the unimportant variables (Kursa and Rudnicki 2010). After Random Forest Boruta sorted out unimportant variables, Random Forest RFE made ranks of important variables (Zhou et al. 2014). Ran-

dom Forest RFE produced the prediction indices, such as the root mean of the squared error (RMSE) and the mean of absolute error (MAE) by adding variables from the most-ranked variables to the least-ranked variables. Finally, Random Forest RFE showed when an additional variable did not have marginal improvement of predicting indices.

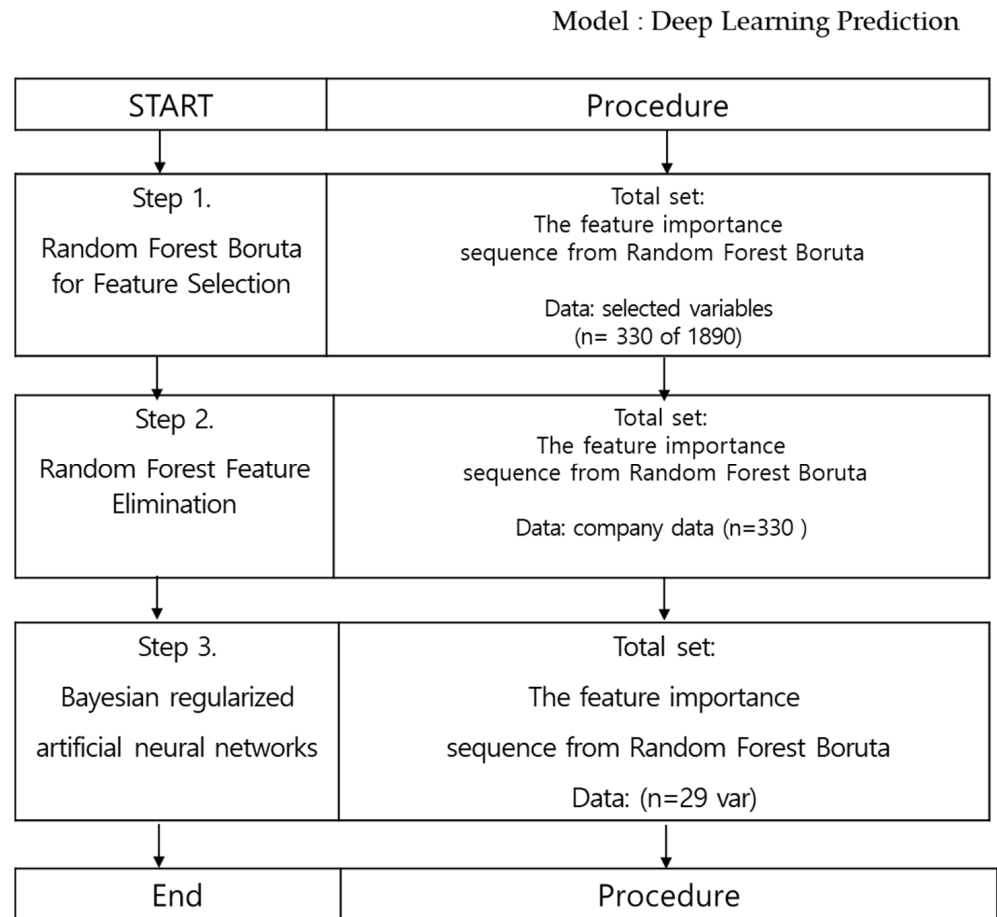


Figure 2. Analytic Algorithm Methodology. Source: Authors’ construction.

In this study, R-studio was utilized. To perform RFE and BRNN, a CARET package (Retrieved 13 March 2021, from <https://cran.r-project.org/web/packages/caret/caret.pdf>.) was utilized, and, to perform Random Forest Boruta, a Boruta package (Retrieved 13 March 2021, from <https://www.analyticsvidhya.com/blog/2016/03/select-important-variables-boruta-package/>) was utilized.

As shown in Table 1, BRNN was utilized as the validating Machine learning method for checking the predicting performance using the important variable list selected from the Random Forest algorithms. To check whether the Machine learning method (i.e., BRNN) was better at predicting the RBC ratio in the next quarter, a conventional linear estimation (i.e., OLS) was performed to predict the RBC ratio in the next quarter. The prediction performance of OLS was compared to BRNN’s prediction performance. The comparison methods used were RMSE and MAE, which were shown as in the below Functions (7) and (8):

$$RMSE = [\Sigma(e_{ti}^2)/(n - 1)]^{1/2} \tag{8}$$

$$MAE = \Sigma(|e_{ti}|)/(n - 1) \tag{9}$$

where, e_{ti} denotes the forecast errors that were calculated in the difference between the observed RBC in the quarter, t , and the predicted RBC in the quarter, $t - 1$. To evaluate the model’s forecasting and performance, RMSE and MAE were generally utilized (Hyndman

and Koehler 2006; Wooldridge 2013). As shown in Functions (2) and (3) above, RMSE and MAE were based on the forecast error (e_{ti}). Therefore, a lower score in RMSE and MAE indicated better performance of models.

6. Results

6.1. Initial Selection by Random Forest Algorithms

From 1890 predictors, 330 predictors were confirmed as important predictors for RBC in the next quarter, including RBC ratio_{t-1}, total shareholders' equity (%), total liabilities (%), policy reserve (%), insurance contract liabilities (%), small- and medium-sized enterprise (SME) loans (%), general accounts with available-for-sale (AFS) security investments (%), non-financial assets (%), claims paid to the group, premium reserves for insurance (%), large corporate loans (%), household loans secured by real estate (%), SME loans (Won), allowances for loan losses by loan type of SME loans, general account refunds for claims paid to the group, annual premiums as a type of premium, financial liabilities by amortized cost, and so on.

Using the 330 predictors that were selected by Random Forest Boruta, Random Forest RFE refined the list with efficient prediction performance. As a result, among 330 predictors, 29 predictors consistently reduced the RMSE and MAE (See Figure 1). After the 29th predictor, the prediction performance fluctuated, and there were some points with increasing RMSE and MAE values.

As shown in Figure 3, the predictor error decreased when adding additional predictors. However, at a certain point (i.e., around the 30th predictor), the tendency no longer showed a notable decrease. As shown in Table 2, the importance weight decreased, but the RMSE and MAE increased by some points at the 30th, 31st, 33rd, 35th, 36th, and 37th predictors. These results can be compared with those in Table 3, which includes the first 29 predictors. The first 29 predictors only showed the decreasing tendencies of RMSE and MAE.

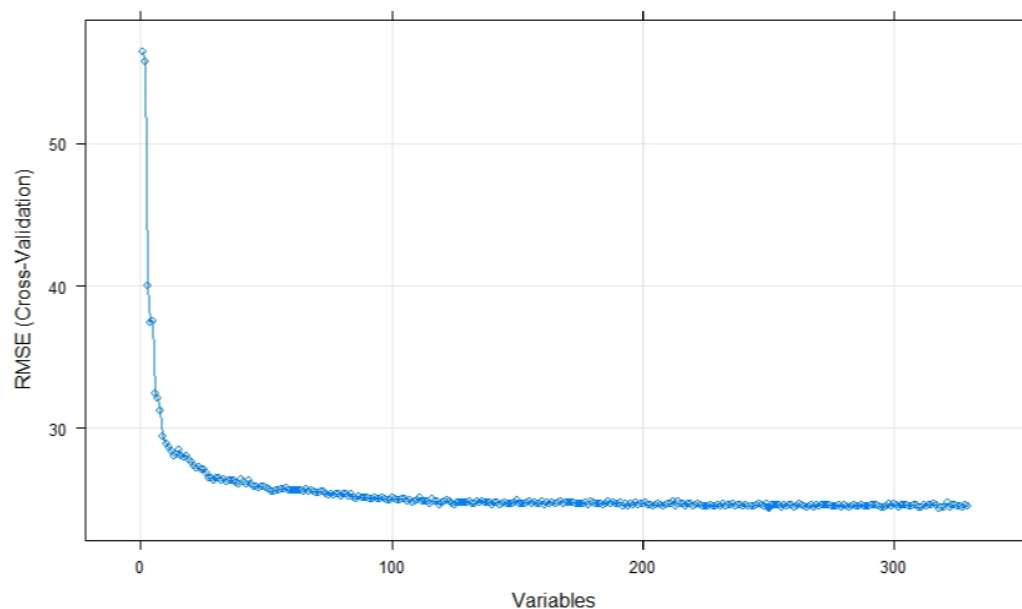


Figure 3. RMSE Changes as a Result of Adding Additional Ranked Predictors. Source: Authors' calculation in R studio.

Table 2. RMSE and MAE after the 29th Predictor.

Predictor	Importance Weight	RMSE	RMSE-SD	MAE	MAE-SD
...					
29th	5.772043847	26.37062	14.44612	20.29887	10.1337
30th	5.768287516	26.50454	14.71273	20.42404	10.31348
31st	5.768078547	26.50908	14.86634	20.34953	10.36669
32nd	5.758095568	26.29367	14.52074	20.34077	10.12503
33rd	5.727512575	26.42472	14.71714	20.36523	10.40581
34th	5.64832017	26.23538	14.66723	20.20653	10.1146
35th	5.629979467	26.30966	14.56628	20.18888	10.18446
36th	5.623392581	26.32962	14.52406	20.21582	10.08676
37th	5.620361217	26.33775	14.52999	20.2968	10.0167
38th	5.586102115	26.1428	14.08058	20.14032	9.734792
39th	5.576443985	26.0594	14.04768	20.03769	9.698879
...

Source: Authors' calculation in R studio.

Table 3. 29 Important Predictors According to Random Forest RFE.

Predictor	Importance Weight	RMSE	RMSE-SD	MAE	MAE-SD
Total Liabilities (%)	16.45244	56.4234	22.54245	43.99274	17.07351
Total Shareholders' Equity (%)	14.13116	55.73478	22.5393	43.60674	17.28434
Total Business Expenses (%)	11.1471	39.9996	15.41544	31.90821	12.04979
Other Liabilities: Bond (%)	8.693667	37.45811	14.41745	29.68432	10.25669
Other Liabilities: Subordinated Bonds (%)	8.347688	37.55018	14.58506	29.34329	10.55252
SME loans (%)	8.288404	32.43499	13.98574	25.43936	10.47889
Financial Liabilities by Amortized Cost (%)	8.248773	32.08482	13.79764	24.95074	10.28786
General Account (AFS) in Security Investment (%)	7.846255	31.24047	13.55237	24.375	10.13581
Insurance Contract Liabilities (%)	7.447804	29.39384	13.43166	22.84536	10.09776
Total Delinquent Loans (%)	7.223881	28.90715	13.44656	22.52241	10.07667
Policy Reserve (%)	7.109629	28.62409	13.50772	22.30659	10.03409
Rate of Return on Asset Investment (%)	6.927131	28.37774	13.69526	22.1116	10.0417
Interest on AFS Securities	6.857443	28.06829	13.78292	21.82381	10.16682
Unpaid Claims	6.468006	28.12855	13.58497	21.93449	9.958229
Risky Asset Ratios in Asset Soundness (%)	6.426568	28.47853	14.36429	21.94059	10.22223
Separate Account (AFS) in Security Investment (%)	6.35893	28.02763	14.08009	21.70934	10.09896
Interest Expenses	6.221578	27.92762	14.1979	21.51127	10.20885
New Account of Individuals	6.182136	28.02603	14.3182	21.5385	10.11705
Security Investment (%)	6.157546	27.80238	14.5411	21.36939	10.26263
Total Number of Agencies	6.035765	27.61104	14.38221	21.29919	10.19745
Overseas Security in General Accounts (AFS)	6.031379	27.39827	14.47423	21.16015	10.31942
Undivided Profits	5.989613	27.20467	14.34592	20.94272	10.25729
Cash and Deposits	5.935971	27.29702	14.31042	21.05666	10.22746
Allowance for Severance and Retirement Benefits	5.890511	27.10333	14.41025	20.8158	10.03827
Number of Stocks Issued	5.849173	27.06379	14.35038	20.84425	10.2582
Total Claims Paid to Groups	5.823229	26.83214	14.44948	20.79202	10.26596
Other Liabilities: Restoration Provision	5.794277	26.52939	14.34872	20.53318	10.11985
Accumulated Other Comprehensive Income	5.780939	26.47633	14.21456	20.48233	10.01474
Interest from Security Investment	5.772044	26.37062	14.44612	20.29887	10.1337

Source: Authors' calculations in R studio.

6.2. Prediction Confirmation Using 29 Predictors: BRNN

Finally, in the case of the 29 predictors, the selection of predictors was validated using BRNN. As explained above, BRNN is one of the Machine learning techniques that minimizes the over-training and overfitting of the model. Therefore, prediction using BRNN can be a conservatively optimal Machine learning technique to be compared with conventional linear estimation (OLS). Below, Table 4 shows the comparison of prediction errors between BRNN and OLS. As shown in Table 4, both RMSE and MAE are lower in

BRNN. In BRNN predictions, the RMSE is 41.15, which is 89.73% of the RMSE shown by OLS (45.86). This result implies that BRNN can perform predictions better than the OLS. However, because RMSE can be exaggerated by outliers (Armstrong 2001), MAE should be used as an alternative method to check the performance (Willmott and Matsuura 2005). As a result, the BRNN model with the selected 29 predictors performed the better prediction. A further implication of 29 predictors will be reviewed in the next discussion.

Table 4. Error Indicators (RMSE and MAE) between BRNN and OLS.

	BRNN		OLS		Comparison
	Mean	SD	Mean	SD	<i>t</i>
RMSE	41.15	-	45.86	-	-
MAE	30.19	20.02	34.54	30.16	-2.11 *

Note. * $p < 0.05$. Source: Authors' calculations in R studio.

All models were resampled 50 times and repeated 9 times. A 50-folds resampling procedure per each iteration was used as a cross-validation method in the study. Table 5 shows the results of prediction using the calculated weight and OLS coefficient with error, but all regression coefficients with 450 times iterations are excluded.

Table 5. Error Indicators (RMSE and MAE) between BRNN and OLS.

		OLS	BRNN			OLS	BRNN
Iteration 1	RMSE	40.79	39.19	Iteration 6	RMSE	44.12	41.90
	MAE	33.09	31.09		MAE	34.96	33.02
	R2	0.74	0.79		R2	0.76	0.76
	Neuron#		2		Neuron#		2
Iteration 2	RMSE	41.71	39.98	Iteration 7	RMSE	43.42	38.54
	MAE	34.54	32.20		MAE	35.26	30.49
	R2	0.73	0.75		R2	0.74	0.80
	Neuron#		2		Neuron#		3
Iteration 3	RMSE	43.45	37.90	Iteration 8	RMSE	40.47	36.51
	MAE	35.32	30.10		MAE	32.67	29.48
	R2	0.75	0.81		R2	0.74	0.80
	Neuron#		2		Neuron#		3
Iteration 4	RMSE	42.32	36.02	Iteration 9	RMSE	42.68	38.30
	MAE	34.41	28.18		MAE	35.01	30.74
	R2	0.76	0.83		R2	0.76	0.80
	Neuron#		3		Neuron#		2
Iteration 5	RMSE	41.99	34.54				
	MAE	33.89	28.00				
	R2	0.72	0.80				
	Neuron#		2				

Source: Authors' calculations in R studio.

7. Conclusions and Discussion

This study aims to find out what the most important indicator for Korean life insurance companies is, in order to identify, evaluate, and manage risks in advance for long-term financial stability through the self-management necessary for corporate social responsibility with financial sustainability.

As introduced above, the financial product of life insurance is a critical product that allows survivors to continue a stabilized financial life after a breadwinner's death (Heo 2020). Therefore, it is important for related people (i.e., insurance policy holders, financial educators, financial policy makers, and the insurers themselves) to predict the stability of life insurance companies for the subsequent time, such as the following quarter. The representative method used to evaluate the stability of life insurance companies was

identifying the RBC ratio, specifically in South Korean life insurance agencies ([Financial Supervisory Service 2017](#)). In this study, a set of predictors for forecasting the next period's RBC ratio for life insurance companies was found using a combination of Machine learning techniques, i.e., Random Forest algorithms and BRNN.

7.1. Methodology Implication

First, compared to the conventional statistical method (OLS), the combination of Random Forest algorithms and BRNN showed better performance in predicting RBC ratios for the next period. This result supports the existing literature ([Bosarge 1993](#); [Heo 2020](#); [Heo et al. 2020](#); [Kudyba 2014](#); [Linoff and Berry 2011](#); [Thompson 2014](#); [Ye 2014](#)), which found that Machine learning techniques can be an alternative method to find better performance in predicting certain outputs.

Considering that the RBC ratio is a solvency screen or early warning for unstable life insurance companies ([Grace et al. 1998](#)), stakeholders such as policy holders and investors can look over the set of predictors for the RBC ratio (see [Table 3](#)) in order to make better decisions when purchasing life insurance policies or in their investments in life insurance companies. Specifically, beyond the main function of securing financial wealth after the loss of a breadwinner, life insurance products may have some investment options. For instance, those who wish to secure future wealth (i.e., retirement planning) may purchase universal or variable life insurance ([Grable 2016](#)). Furthermore, life insurance can be utilized as a type of investment, such as tax savings on estates and bequests ([Cymbal 2013](#); [Kait 2012](#); [Whitelaw 2014](#)). However, if a life insurance company is not stable, with a low RBC ratio, then retirement planning or investment planning with life insurance options fails to support an individual's purchasing purpose. Therefore, these signals (i.e., 29 predictors) may help Korean consumers of life insurance to discover which life insurance companies demonstrate good signals or bad signals for the next period's RBC ratio.

7.2. Industrial Implications

In addition, as shown below in [Table 6](#), the set of important predictors can be cataloged into three categories: liabilities and expenses, other financial predictors, and predictors from business performance. The first two categories (liabilities and expenses and other financial predictors) are obviously natural since the RBC ratio is calculated from those pieces of financial information (see [Function \(1\)](#) and [Figure 1](#)).

However, there are some predictors that do not fall into the categories of financial predictors. For instance, sincere business performance as part of an insurance company's obligations (i.e., unpaid claims, undivided profit, and total claims paid to groups) was found to be amongst the predictors. This factor implies that how ethically a life insurance company performs business is an important component in their RBC ratio. Considering that the RBC ratio is calculated from various types of risks ([Financial Supervisory Service 2017](#)), it may be natural that any ethical business performance may contribute to the next period's RBC ratio as an indicator of a positive characteristic for a life insurance company; on the other hand, unethical business will be related to negative characteristics. Another interesting finding from [Table 5](#) is the total number of agencies for whom the allowance for severance and retirement benefits was found to be an important indicator for the RBC ratio. Those predictors were employee benefits inside of a life insurance company. This finding implies that how a life insurance company treats its employees is important in their RBC ratio.

Table 6. Three Categories of Predictors found by Machine Learning Techniques.

Liabilities and Expenses	Other Financial Predictors	Predictors from Business Performance
	Total Shareholders' Equity	
	SME Loans	
	General Account (AFS) in Security Investment	Unpaid Claims
Total Liabilities		New Account of Individuals
Total Business Expenses	Total Delinquent Loans	Total Number of Agencies
Other Liabilities from Bond	Policy Reserve	Undivided Profit
Other Liabilities from Subordinated Bond	Rate of Return on Asset Investment	Cash and Deposits
Financial Liabilities by Amortized Cost	Interest on AFS Securities	Allowance for Severance and Retirement Benefits
Insurance Contract Liabilities	Risky Asset Ratio in Asset Soundness	Number of Stock Issues
Interest Expenses	Separate Account (AFS) in Security Investment	Total Claims Paid to Groups
Other Liabilities from Restoration Provision	Total Investment	Accumulated Other Comprehensive Income
	Overseas Security in General Account (AFS)	
	Interest from Security Investment	

Note. AFS means available for sale. Source: Authors' construction.

7.3. Limitations

The main purpose of this study was to find the most important set of variables to monitor life insurance companies' ability to pay and better predict future RBC rates. In this study, a critical predictor was defined as being one that would predict the RBC ratio for the next period. However, as explained in the introduction, the explanatory power of each factor was intentionally ignored in this study due to the predictive methodology. Therefore, future studies should investigate newly discovered difficulties such as unpaid payments, the total number of machines, unpaid dividend profits, severance pay and severance pay payments, and total group payments. Sustainable insurance emphasizes risk management for financial stability. To this end, this study analyzed all 1891 strains used to supervise domestic life insurance companies. This result shows that a series of predictors used to predict the RBC ratio over the next period will contribute to being able to better identify, manage, and perform the risks of sustainable insurance with financial soundness. In other words, since this study mainly focuses on finding a set of variables to better predict RBC ratios, this study limits the purpose of variable selection instead of comparing various Machine learning techniques with the usefulness of specific algorithms in terms of predicting outputs. Based on the selected variables, other researchers can investigate external factors in the process of calculating the RBC ratio and view it as a possible research topic for future research to improve financial information.

Author Contributions: J.P.—Conceptualization, data management, original draft preparation, methodology, investigation, formal analysis, writing; M.S.—Project administration, review and editing, writing revision, supervision, validation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Details of RBC ratio calculation components.

Type	Life Insurance	Key Categories	Non-Life Insurance
Available Capital	① Summed Items (Core capital + Supplementary Capital) – ② Deducted Items + ③ Subsidiary-related items (subtraction and summation)		
Core Capital	<ul style="list-style-type: none"> “Paid-in Capital” (excluding “cumulative preferred stock and hybrid bond”) “Capital surplus” (excluding “cumulative preferred stock and hybrid bond”) Retained earnings (excluding Reserve for Credit Losses) Accumulated other comprehensive income Less than 25% of the equity capital of the issued amount of “Hybrid bonds” (i.e., the equity capital under Article 2-15 of the Insurance Business Act based on IFRS consolidated financial statements) Non-controlling interests in the above five categories 		
Summed Items	<ul style="list-style-type: none"> The amount calculated by subtracting the surrender charge from the “net premium-type reserve” and additional reserve regarding Liability Adequacy test The reserves to be refunded upon termination of the savings-type insurance and the reserve in excess of additional accumulated amount related to the Liability Adequacy test 		
Supplementary Capital	<ul style="list-style-type: none"> “Loan Loss Provision and Reserve for Credit Losses for assets classified as normal and precautionary” The amount exceeding 25% of total Shareholders’ equity of issued “Redeemable preferred stock”, “Repayment of subordinated debt, Total amount of hybrid bond” and “cumulative preferred stock issuance amounts” (limited to subtracting 20/100 every year if the remaining maturity of “subordinated debt” and repaid preferred stock is within 5 years) Non-controlling interest in the above two categories of subordinated insurance company. “Reserves for policyholders’ profit dividends, Reserve for compensation for losses on dividend-paying” “Reserve for policyholder’s profit dividend stabilization” “Gain(Loss) on valuation of Available-for-Sale financing assets”, “Policyholder’s equity adjustment and revaluation Surplus” “Deferred tax liabilities related to emergency risk reserves (only available for “Deferred income tax” on the book calculated by insurance company 		
Deduction	<ul style="list-style-type: none"> “Intangible assets”, the marketability of which is difficult to measure, such as unrevised contract costs and goodwill “Prepaid cost” without marketability, “Deferred tax asset”, and expected cash dividends The amount exceeding the fair value amount on the book value of equity investments (excluding related subsidiaries, non-insurance financial companies, and overseas subsidiaries that are not included in the consolidated scope of the RBC consolidated financial statements) Accumulated and unrealized valuation gains and losses of financial liabilities due to changes in credit risk Amount of Reserve for Credit Losses for asset under fixed amount that cannot be accumulated in capital due to untreated losses “Discounts on stock issuance”, “treasury stocks” (including treasury stocks invested by private equity firms), changes in equity method accounting “Capital raising means” mutually held by other insurance companies for the purpose of improving the Solvency Margin Ratio 		
Subsidiary-related items	<p>< Deducted items based on the Consolidated financial statement ></p> <ul style="list-style-type: none"> The amount of investment in overseas subsidiaries that have not secured the consistency, sufficiency, and objectivity of data to calculate Standard Amount of Solvency Margin Investment amount of domestic- and overseas-related insurance companies Investment amount of domestic- and overseas-related non- insurance companies <p>< Summed items based on the Consolidated financial statement ></p> <ul style="list-style-type: none"> Equivalent amount of equity ratio of the (insurance) parent company in the available capital based on the RBC of domestic affiliated company. Equivalent amount of equity ratio of the (insurance) parent company in the available capital based on the RBC of the domestic subsidiary and affiliated company. Equivalent amount of equity ratio of the (insurance) parent company in the available capital based on the RBC of the overseas affiliated company with consistency, sufficiency, and objectivity of the data used to calculate Standard Amount of Solvency Margin Equivalent amount of equity ratio of the (insurance) parent company in the available capital based on the RBC of overseas-subordinated, affiliated, and non- insurance companies with consistency, sufficiency, and objectivity of the data used to calculate Standard Amount of Solvency Margin 		

Source: [Financial Supervisory Service \(2017\)](#).

Table A2. Risk Coefficient for RBC ratio.

Groups	Insurance	Interest Rate	Credit	Market
Insurance	1	0.25	0.25	0.25
Interest rate	0.25	1	0.5	0.5
Credit	0.25	0.5	1	0.5
Market	0.25	0.5	0.5	1

Source: Korea Ministry of Government Legislation, Article 5-7-3 (2020).

Notes

- ¹ Solvency margin denotes the capital, reserve for dividends to policyholders, allowance for non-performing loans, subordinated loans, deferred acquisition costs, goodwill, and other similar amounts determined and announced by the Financial Services Commission (Financial Supervisory Service 2017). This indicates the amount remaining after deducting other amounts determined and announced by the Financial Services Commission.
- ² The term solvency margin ratio means the ratio obtained by dividing the amount of solvency margin by the standard amount of solvency margin. The solvency margin ratio must be maintained at no less than 100/100. The term standard amount of solvency margin means the results produced by converting any risks incurred while running an insurance business into an amount of money using the methods determined and publicized by the Financial Services Commission.

References

- Armstrong, Jon Scott. 2001. *Principles of Forecasting: A Handbook for Researchers and Practitioners (International Series in Operations Research & Management Science)*. Boston: Springer.
- Barboza, Flavio, Herbert Kimura, and Edward Altman. 2017. Machine Learning Models and Bankruptcy Prediction. *Expert Systems with Applications* 83: 405–17. [CrossRef]
- Beutel, Johannes, Sophia List, and Gregor von Schweinitz. 2019. Does Machine Learning Help Us Predict Banking Crises? *Journal of Financial Stability* 45: 100693. [CrossRef]
- Biau, Gérard. 2012. Analysis of Random Forest model. *Journal of Machine Learning Research* 13: 1063–95.
- Bosarge, W. E. 1993. Adaptive processes to exploit the nonlinear structure of financial markets. In *Neural Networks in Finance and Investing*. Edited by Robert R. Trippi and Efraim Turban. Chicago: Probus Publishing Company, pp. 371–402.
- Breiman, Leo. 2001. Random Forest. *Machine Learning* 45: 5–32. [CrossRef]
- Brunton, Steven L., and J. Nathan Kutz. 2019. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge: Cambridge University Press.
- Burden, Frank, and Dave Winker. 2008. Bayesian regularization of neural network. In *Artificial neural Networks. Methods in Molecular BiologyTM*, 458. Edited by David J. Livingstone. Totowa: Humana Press, pp. 23–42.
- CBS. 2008. Insurance industry floundered in 2008. *CBS News*. Available online: <https://www.cbsnews.com/news/insurance-industry-floundered-in-2008/> (accessed on 23 September 2021).
- Chatfield, Chris, and Haipeng Xing. 2019. *The Analysis of Time Series*. London: Chapman and Hall/CRC.
- Cymbal, Kenneth M. 2013. Choosing a family member as trustee of an irrevocable life insurance trust. *Journal of Financial Services Professionals* 67: 41–52.
- Ding, Kexing, Baruch Lev, Xuan Peng, Ting Sun, and Miklos A. Vasarhelyi. 2020. Machine Learning Improves Accounting Estimates: Evidence from Insurance Payments. *Accounting Technology & Information Systems eJournal* 25: 1098–134.
- Enforcement Decree of the Insurance Business Act § 65. 2019. Available online: <https://law.go.kr/LSW/lsInfoP.do?ancYnChk=undefined&efYd=&lsiSeq=210659#0000> (accessed on 20 November 2020).
- Enforcement Decree of the Insurance Business Act § 65-2. 2019. Available online: <https://law.go.kr/LSW/lsInfoP.do?ancYnChk=undefined&efYd=&lsiSeq=210659#0000> (accessed on 20 November 2020).
- Financial Supervisory Service. 2017. *Guide to Korea's Risk-Based Capital for Korean Insurance Companies*. Available online: [https://www.fss.or.kr/fss/kr/bbs/view.jsp?page=2&url=/fss/kr/1240186854180&bbsid=1240186854180&idx=1485240365199&num=28&stitle=%BA%B8%C7%E8%C8%B8%BB%E7%20RBC%C1%A6%B5%B5%20%C7%D8%BC%B3%BC%AD\(2017.1%BF%F9\)](https://www.fss.or.kr/fss/kr/bbs/view.jsp?page=2&url=/fss/kr/1240186854180&bbsid=1240186854180&idx=1485240365199&num=28&stitle=%BA%B8%C7%E8%C8%B8%BB%E7%20RBC%C1%A6%B5%B5%20%C7%D8%BC%B3%BC%AD(2017.1%BF%F9)) (accessed on 20 November 2020).
- Financial Supervisory Service. n.d. History. Available online: <http://english.fss.or.kr/fss/eng/wpge/eng111.jsp> (accessed on 20 November 2020).
- Genuer, Robin, Jean-Michel Poggi, and Christine Tuleau-Malot. 2010. Variable selection using random forests. *Pattern Recognition Letters* 31: 2225–36. [CrossRef]
- Grable, John E. 2016. *The Case of Approach to Financial Planning: Bridging the Gap between Theory and Practice*, 3rd ed. Erlanger: National Underwriter Company.
- Grace, Martin F., Scott E. Harrington, and Robert W. Klein. 1998. Risk-Based Capital and Solvency Screening in Property-Liability Insurance: Hypotheses and Empirical Tests. *Journal of Risk and Insurance* 65: 213. [CrossRef]
- Guyon, Isabelle, Jason Weston, Stephen D. Barnhill, and Vladimir Naumovich Vapnik. 2004. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 46: 389–422. [CrossRef]
- Heo, Wookjae. 2020. *The Demand for Life Insurance: Dynamic Ecological Systemic Theory Using Machine Learning Techniques*. Cham: Springer.
- Heo, Wookjae, Jae Min Lee, Narang Park, and John E. Grable. 2020. Using Artificial Neural Network techniques to improve the description and prediction of household financial ratios. *Journal of Behavioral and Experimental Finance* 25: 100273. [CrossRef]
- Hsiao, Shu-Hua, and Thou-Jen Whang. 2009. A study of financial insolvency prediction model for life insurers. *Expert Systems with Applications* 36: 6100–7. [CrossRef]
- Hutagaol, B. Junedi, and Tuga Mauritsius. 2020. Risk Level Prediction Of Life Insurance Applicant Using Machine Learning. *International Journal of Advanced Trends in Computer Science and Engineering* 9: 2213–20. [CrossRef]

- Hyndman, Rob J., and Anne B. Koehler. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22: 679–88. [CrossRef]
- Insurance Business Act § 123. 2019. Available online: <https://law.go.kr/LSW/lsInfoP.do?ancYnChk=undefined&efYd=&lsiSeq=206443#0000> (accessed on 22 November 2020).
- Insurance Business Supervision Regulation § 2-6-3. 2020. Available online: <https://law.go.kr/LSW/admRulInfoP.do?admRulSeq=2100000190369> (accessed on 22 November 2020).
- Insurance Business Supervision Regulation § 7-1. 2020. Available online: <https://law.go.kr/LSW/admRulInfoP.do?admRulSeq=210000190369> (accessed on 22 November 2020).
- Insurance Business Supervision Regulation § 7-14. 2020. Available online: <https://law.go.kr/LSW/admRulInfoP.do?admRulSeq=210000190369> (accessed on 22 November 2020).
- Insurance Business Supervision Regulation § 7-17. 2020. Available online: <https://law.go.kr/LSW/admRulInfoP.do?admRulSeq=210000190369> (accessed on 22 November 2020).
- Insurance Business Supervision Regulation § 7-18. 2020. Available online: <https://law.go.kr/LSW/admRulInfoP.do?admRulSeq=210000190369> (accessed on 22 November 2020).
- Insurance Business Supervision Regulation § 7-19. 2020. Available online: <https://law.go.kr/LSW/admRulInfoP.do?admRulSeq=210000190369> (accessed on 22 November 2020).
- Insurance Business Supervision Regulation § 7-2. 2020. Available online: <https://law.go.kr/LSW/admRulInfoP.do?admRulSeq=210000190369> (accessed on 22 November 2020).
- Insurance Business Supervision Regulation § 7-36. 2020. Available online: <https://law.go.kr/LSW/admRulInfoP.do?admRulSeq=210000190369> (accessed on 22 November 2020).
- International Association of Insurance Supervisor. 2021. Insurance Core Principles and ComFrame. Available online: <https://www.iaisweb.org/page/supervisory-material/insurance-core-principles-and-comframe> (accessed on 22 January 2021).
- Kait, R. E. 2012. One life insurance size doesn't fit all estate planning situations. *Journal of Financial Planning* 26: 38–39.
- Kartasheva, Anastasia V., and Mikhail Traskin. 2013. *Insurers' Insolvency Prediction Using Random Forest Classification*. Rochester: SSRN.
- Kohavi, Ron, and George H. John. 1997. Wrappers for Feature Subset Selection. *Artificial Intelligence* 97: 273–324. [CrossRef]
- Kou, Gang, Xiangrui Chao, Yi Peng, Fawaz E. Alsaadi, and Enrique Herrera-Viedma. 2019. Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy* 25: 716–42. [CrossRef]
- Kudyba, Stephan, ed. 2014. *Big Data, Mining, and Analytics*. New York: CRC Press and Taylor and Francis.
- Kursa, Miron Bartosz, and Witold R. Rudnicki. 2010. Feature Selection with the Boruta Package. *Journal of Statistical Software* 36: 1–13. [CrossRef]
- Lau, King Tong, Weimin Guo, Breda M. Kiernan, Conor Slater, and Dermot Diamond. 2009. Non-linear carbon dioxide determination using infrared gas sensors and neural networks with Bayesian regularization. *Sensors and Actuators B-Chemical* 136: 242–47. [CrossRef]
- Linoff, Gordon S., and Michael J. A. Berry. 2011. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 3rd ed. Indianapolis: Wiley.
- Liu, Li Xian, Shuangzhe Liu, and Milind Sathye. 2021. Predicting Bank Failures: A Synthesis of Literature and Directions for Future Research. *Journal of Risk and Financial Management* 14: 474. [CrossRef]
- Muttill, Nitin, and Kwok-wing Chau. 2007. Machine-learning paradigms for selecting ecologically significant input variables. *Engineering Applications of Artificial Intelligence* 20: 735–44. [CrossRef]
- National Association of Insurance Commissioners. 2020. *Risk-Based Capital*. Available online: https://content.naic.org/cipr_topics/topic_riskbased_capital.htm (accessed on 15 March 2021).
- National Association of Insurance Commissioners. n.d. Post-Crisis Financial System Reform: Impact on the U.S. Insurance Industry in the Evolving Regulatory Landscape. Available online: https://www.naic.org/capital_markets_archive/170120.htm (accessed on 15 March 2021).
- Petropoulos, Anastasios, Vasilis Siakoulis, Evangelos Stavroulakis, and Nikolaos E. Vlachogiannakis. 2020. Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting* 36: 1092–113. [CrossRef]
- Rejda, George E. 2008. *Principles of Risk Management and Insurance*, 10th ed. Boston: Pearson/Addison Wesley.
- Rudin, Cynthia. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1: 206–15. [CrossRef]
- Sariev, Eduard, and Guido Germano. 2019. Bayesian regularized artificial neural networks for the estimation of the probability of default. *Quantitative Finance* 20: 311–28. [CrossRef]
- Serrano-Cinca, Carlos, and Begoña Gutiérrez-Nieto. 2013. Partial Least Square Discriminant Analysis for bankruptcy prediction. *Decision Support Systems* 54: 1245–55. [CrossRef]
- Thompson, Wayne. 2014. Data mining methods and the rise of big data. In *Big Data, Mining, and Analytics*. Edited by Kudyba Stephan. New York: CRC Press and Taylor and Francis, pp. 71–101.
- Thoyts, Rob. 2010. *Insurance Theory and Practice*. London: Routledge.
- Whitelaw, E. Randolph. 2014. How to relieve the plight of unskilled irrevocable life insurance trust trustees unfamiliar with their duties. *Journal of Financial Service Professionals* 68: 44–49.

-
- Willmott, Cort J., and Kenji Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30: 79–82. [[CrossRef](#)]
- Wooldridge, Jeffrey M. 2013. *Introductory Econometrics: A Modern Approach*, 5th ed. Mason: South Western, Cengage Learning.
- Ye, Nong. 2014. *Data Mining: Theories, Algorithms, and Examples*. Boca Raton: CRC Press and Taylor and Francis.
- Zhou, Qifeng, Haotian Zhou, Qingqing Zhou, Fan Yang, and Linkai Luo. 2014. Structure damage detection based on random forest recursive feature elimination. *Mechanical Systems and Signal Processing* 46: 82–90. [[CrossRef](#)]