

Author

Martin Elliot
Director-Insurance Research

Contact Us

Americas
+1.212.553.1658
clientservices@moodys.com

Europe
+44.20.7772.5454
clientservices.emea@moodys.com

Asia (Excluding Japan)
+85 2 2916 1121
clientservices.asia@moodys.com

Japan
+81 3 5408 4100
clientservices.japan@moodys.com

Using Machine Learning in Actuarial Modeling to improve model efficiency through better clustering

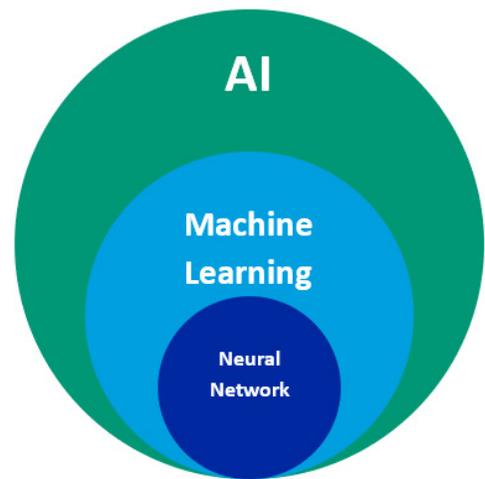
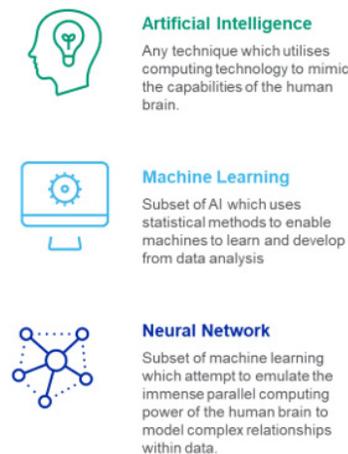
Artificial intelligence (AI) and in particular machine learning are increasingly becoming a popular and important component of an actuary's toolkit. As the amount of data available to actuaries continues to increase, these techniques can allow them to extract more insight; which improves the quality of their analysis, and the resulting decisions the insurer makes.

This paper will explore one of the ways Moody's Analytics has implemented AI within our AXIS™ actuarial system to facilitate the use of powerful model efficiency techniques, and reduce the run time on large models, without adversely affecting accuracy.

What is Machine Learning?

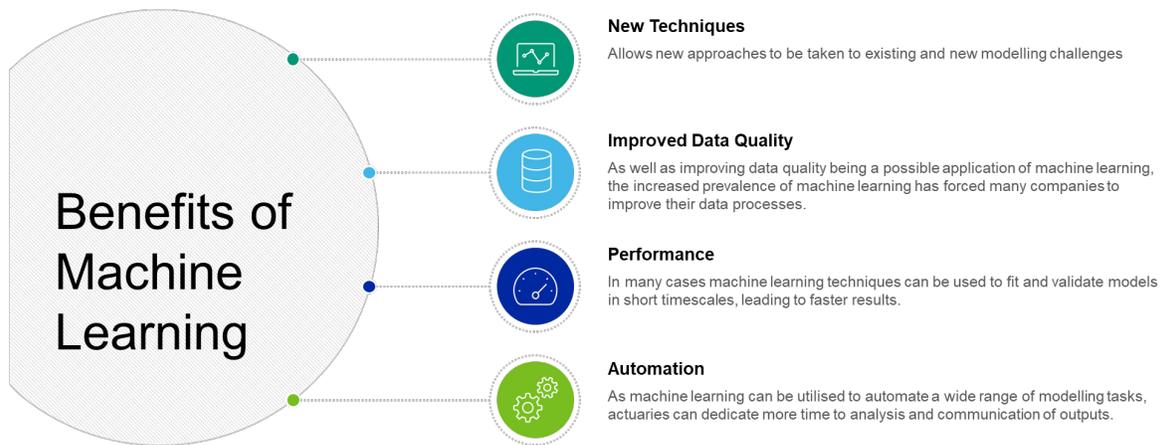
Machine learning is a method of teaching computers to analyze data, learn from it, and then make a determination or prediction regarding new data. Rather than hand-coding a specific set of task-focused instructions, the machine is "trained" using large amounts of data and algorithms to learn the task.

Figure 1 AI Definitions



Machine learning has already brought several benefits for actuaries:

Figure 2 Benefits of Machine Learning in Actuarial Modeling



Machine learning techniques have been applied to a range of actuarial applications including:

- » Mortality forecasting¹
- » Customer support
- » Marketing
- » Predicting future claims as part of pricing tool
- » Utilizing unstructured data in forecasting interest rates²

The Challenge

Under IFRS 17 and other new reporting frameworks, insurers must:

- » Increasingly run actuarial models to model their business on a seriatim (policy by policy) basis
- » Run those models a greater number of times in a reporting cycle

The increasing amount of policy data that insurers have available fuels the challenge - as well as the requirement for better, more timely management information. With these demands, the resulting costs and extended runtimes can be onerous, even with today's powerful technology. Therefore, management's ability to make fast, informed decisions based on timely results may be impaired.

Our Solution - Agglomerative Clustering

¹ <https://www.theactuary.com/features/2020/08/05/machine-learning-deep-end>

² <https://www.actuaries.org.uk/documents/practical-application-machine-learning-within-actuarial-work>

To address this challenge, Moody's Analytics has implemented many powerful model efficiency techniques to help improve performance. In particular, the AXIS system supports both traditional and advanced data compression/grouping methods. These methods reduce data models that contain many policies into fewer model points which each represents several original policies.

Policy grouping traditionally involves using a fixed, predefined set of rules to group individual policies into a smaller number of representative policies. This is done without regard to the varying impact of that compression on model accuracy.

Clustering is a statistical technique commonly used to identify naturally occurring groups within large amounts of data. This technique can be applied to insurance modeling and overcomes the disadvantages of traditional grouping. In a seriatim data model, natural data clusters can be found due to similarities in product features, timing of policy issuance, policy holder age, and fund guarantee levels. All of these features impact the modeled cash flow patterns. Clustering compression techniques in the AXIS system work within these clusters to reduce the impact on accuracy, and enable a flexible approach to determining the final amount of compression to be used.

Like traditional grouping, a "central" policy represents the whole cluster, by scaling up the central policy's calculated results. However, clustering differs from traditional grouping in many significant ways.

Figure 3 Clustering v. Traditional Grouping

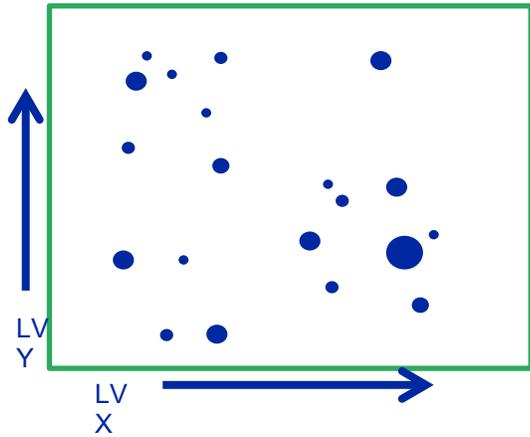
Traditional Grouping	Clustering
Predefined grouping rules	No requirement for predefined groupings. Determined based on the actual data
Compression ratio an output of the process	Target compression ratio can be set as an input to the process
Changing compression ratio requires new grouping rules to be defined and the grouping process to be repeated in full.	The compression ratio can be adjusted to achieve an acceptable distortion when compared to the full seriatim data model
Different data models require creation of multiple predefined groupings and running of multiple processes.	Multiple Data Models of different compression levels can be generated for different purposes in one process

There are various algorithms used in statistical analysis to perform clustering of data. In the AXIS system, an "agglomerative" clustering technique is used. Under this approach:

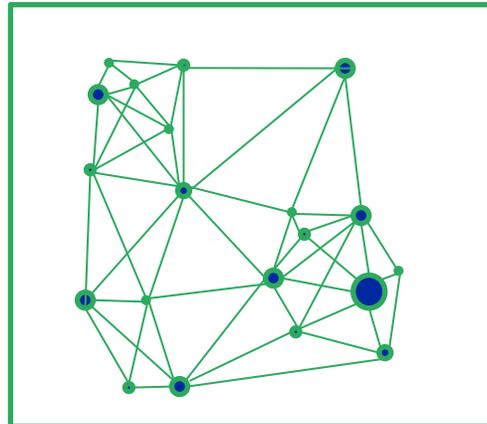
- » Each policy starts as its own "cluster"
- » Distance (similarity) between every pair is calculated using chosen location variables (characteristics) for each policy
- » The least important cluster by size and distance is found and merged into the nearest cluster
- » Clustering process continues until the target compression is achieved.

The following diagrams show the clustering process for a simple example, with 20 seriatim policies of varying sizes and characteristics and a target compression ratio of 20% - reducing the policies into four clusters.

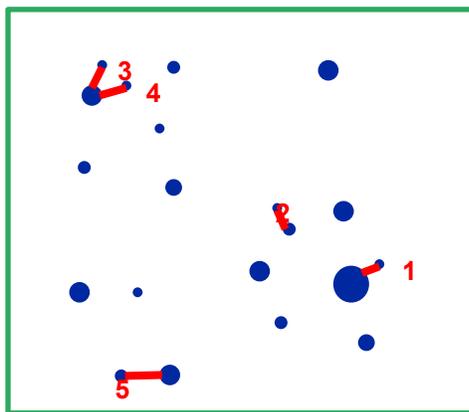
Step 1: Seriatim data assigned location variables and size



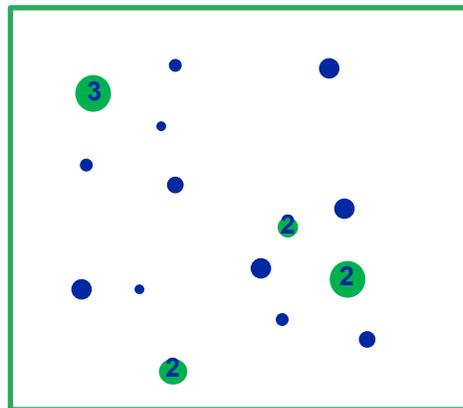
Step 2: Distances between all pairs calculated



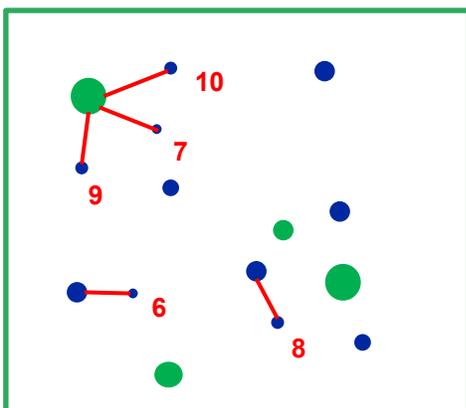
Step 3: Closest pairs identified



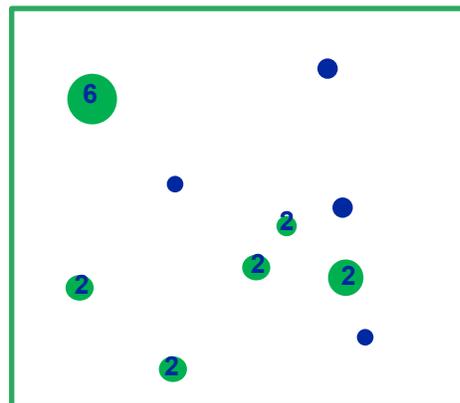
Step 4: Least important clustered with larger adjacent point



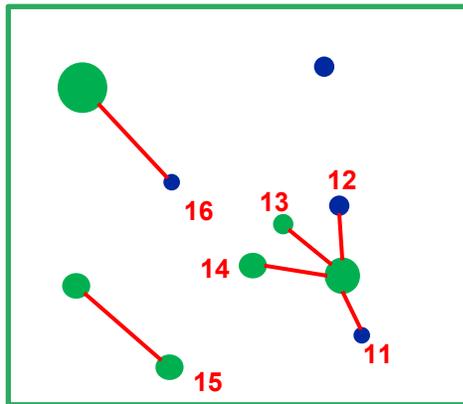
Step 5: Next closest pairs identified



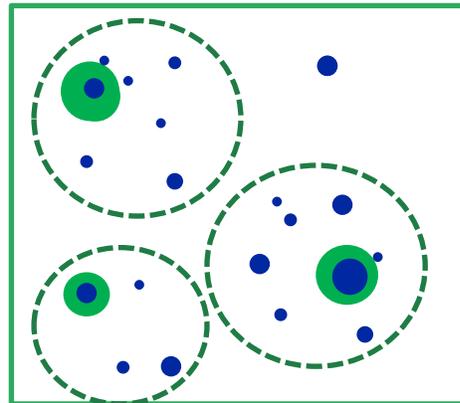
Step 6: Least important clustered with larger adjacent point



Step 7: Remaining closest pairs identified to reach target



Step 8: Final clustering performed. Highly compressed model and original model shown.



Neural Networks - How Moody's Analytics has Incorporated AI into our AXIS solution

Moody's Analytics has extended functionality using machine learning techniques, in particular Neural Networks to improve the performance of the clustering process.

Within the clustering process, the definition of the mapping (or location variables) has a critical impact on the efficiency of the clustering. Ideally, policies from the same cluster should deliver model results close to those of the cluster's representative policy. In this case, the error introduced by the substitution of a policy by its cluster representative will be small, suggesting that the best location variables are the actual policy results themselves (for example, a seriatim reserve). However, this would involve performing the full calculations for each policy, negating the advantage of clustering.

Neural networks are non-linear statistical data modeling or decision-making tools which attempt to emulate the immense parallel computing power of the human brain. Such networks can be used to model complex relationships between data inputs and outputs or to find hidden patterns in data. Neural networks effectively learn from their experience and adapt their logic and predictive models accordingly. Moody's Analytics has developed a technique of creating and optimizing the location variables needed for effective clustering using a neural network approach.

Configuring and using a neural network involves the following steps:

1. Select neural network architecture
2. Select activation functions
3. Train the neural network to find parameters which optimize it
4. Use the neural network to predict output values

Neural network architecture

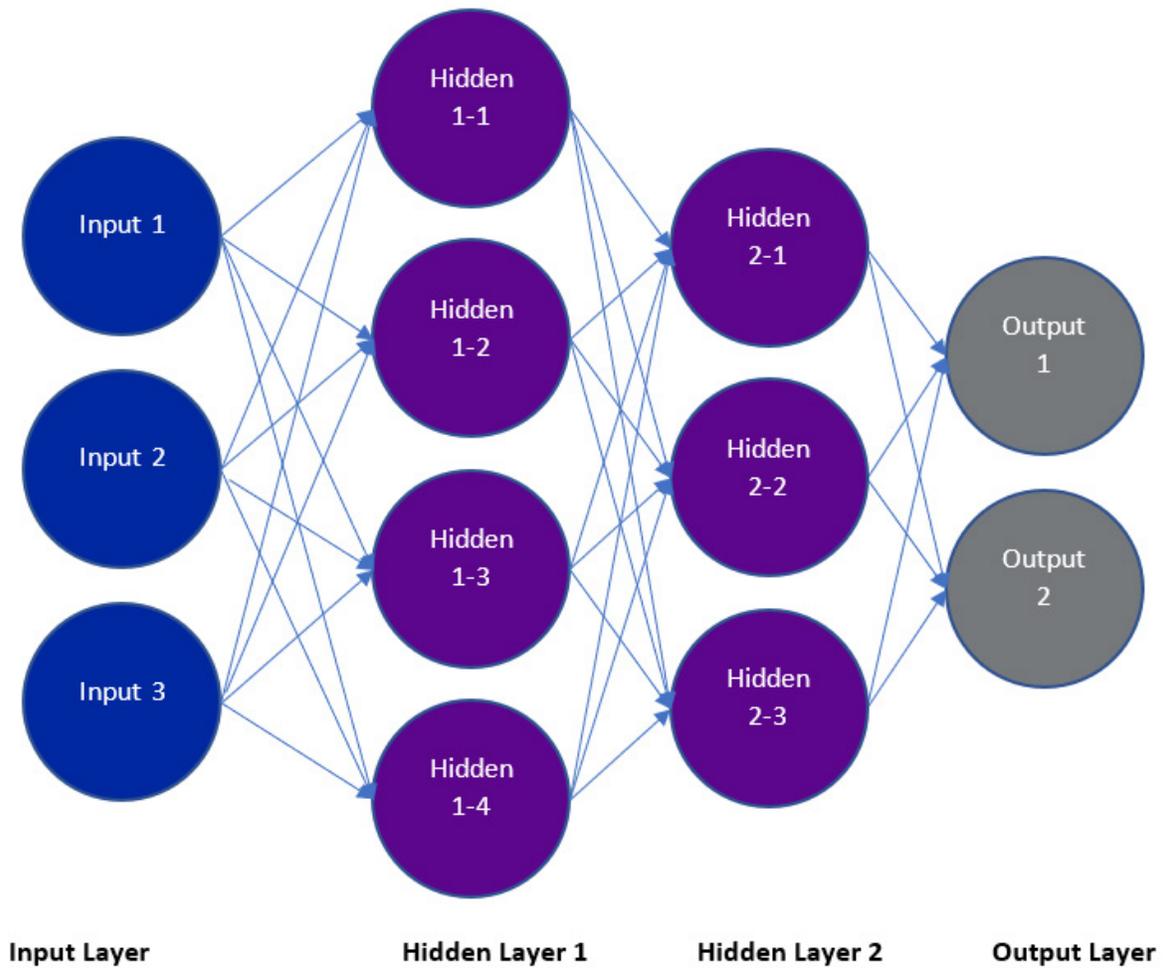
A neural network contains a sequence of layers:

- » Input layers which provide the initial data used within the neural network
- » Output layers which produce given outputs for the neural network to deliver the result.

- » Hidden layers which are intermediate layers between the input and output layers. The number of hidden layers can be user-defined. In theory, one hidden layer is sufficient if it has enough nodes. In practice, deep networks (those with multiple hidden layers) do the same job with fewer total number of weights and nodes. We have found that two or three hidden layers work well.

Each layer is made up of a set of nodes. The number of inputs into and outputs required from the Neural Network defines the number of nodes in the Input and Output layers. Each node in the hidden layers takes as its input, a combination of the outputs from the previous layer. The combination is usually a weighted sum shifted by a bias term. Each node then processes its signal through an activation function and outputs it to the next layer.

Figure 4 Neural Network



In the Neural Network architecture shown in Figure 1, the outputs from each layer are calculated as follows:

For Hidden Layer 1:

$$x_j^1(Z) = \Phi^1 \left(\sum_{i=1}^3 Z_i w_{i,j}^1 + w_{0,j}^1 \right)$$

For Hidden Layer 2:

$$x_j^2(X) = \Phi^2 \left(\sum_{i=1}^4 x_i^1 w_{i,j}^2 + w_{0,j}^2 \right)$$

For the output layer:

$$Y_j(X) = x_j^3(X) = \Phi^3 \left(\sum_{i=1}^3 x_i^2 w_{i,j}^3 + w_{0,j}^3 \right)$$

Where:

- » Z_i is the input to the Neural network from Input node i
- » $w_{i,j}^k$ is the weight applied to the input to node j in layer k coming from node i of the previous layer (with $i=0$ identifying the bias term)
- » $\Phi^k(x)$ is the activation function used to process inputs in layer k
- » $x_j^k(X)$ is the output from node j in layer k

Activation Functions

The optimal choice of the activation function depends on the application. The AXIS system currently supports the following activation functions (depending upon a layer):

$$\text{Sigmoid: } \sigma(z) = \frac{1}{1 + e^{-z}}$$

$$\text{Tanh: } \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

$$\text{Linear: } l(z) = z$$

$$\text{Rectified Linear Unit (ReLU): } \text{relu}(z) = \max(0, z)$$

$$\text{SoftPlus: } s(z) = \ln(1 + e^z)$$

In choosing an activation function to use, clients should consider the relationship between the inputs and outputs of the relevant nodes. Also consider boundaries to be applied to the outputs. Examples of considerations include:

- » If non-linearity is observed in these relationships, then the linear activation function will not be appropriate as it will not capture these patterns. This function also may not be suitable if bounds on the outputs exist as there are no upper or lower bounds on the values that the outputs can take.
- » If a lower bound of 0 is required, then the ReLU function may be appropriate but is still unbounded. It is also non-differentiable at $z=0$. This problem can be avoided by using the SoftPlus function. As z tends to $-\infty$ the Softplus function tends to e^z which approaches 0 and as z tends to ∞ it tends to $z + e^{-z}$ which approaches z so is a smooth approximation to the ReLU function.
- » The Tanh and Sigmoid function will allow non-linear relationships to be captured. It will also bound output values between 0 and 1 (Sigmoid) or -1 and 1 (Tanh)

After the number of layers and activation functions have been selected, the neural network can be used to approximate the desired result for each policy. The process for using the neural network can be split into two sub-processes:

- » Training: Find the parameters (weights and biases) which optimize the neural network.
- » Scoring: Using the neural network to predict the outputs for the full set of policies.

Train the neural network

A subset of the policies can be run through the full model to create outputs for training data. The Neural Network is trained using these results. Given the input and output variable values for the training set; the training process aims to find the weights that minimize the square error across all outputs for all records.

$$Square\ Error = \sum_{i=1}^{\# training\ records} \sum_{j=1}^{\# output\ nodes} (Y_j^{predicted}(Z) - Y_j^{actual}(Z))^2$$

Where:

- » $Y_j^{predicted}(Z)$ are the output values produced by the neural network
- » $Y_j^{actual}(Z)$ are the actual policy results

The training data used should be representative of the full set of policies – both in terms of distribution and number of records. The training will lead to a better approximation of the input values in the areas for which more data is available. To ensure that the network also produces valid outputs for policies outside of those used for training, the training data may be split into two parts. Some policies will not be included in the training process, but scored by the resultant neural network. This means that the errors between the accurate full model results and the scored results can be calculated.

Use the neural network to predict output values

After the Neural Network is trained (optimal weights are found), it can be used to score the outputs for policies not included in the training set. In the clustering process, the complete policy set's Neural Network scoring results are used as location variables. As the results approximate the actual policy results; utilizing them in this way improves the likelihood that the representative policies will generate results, consistent with those of the policies clustered together.

Conclusion

The continued development of machine learning techniques introduces interesting possibilities for how they can be used to extend the usage of actuarial models. Through the adoption of new techniques, improved data quality, greater performance, and increased automation, these techniques allow actuaries to make more effective use of the increasing amounts of data available.

The ability to train neural networks and use them to replace trial and error experimentation within the AXIS actuarial system opens up a range of applications. This paper has highlighted one of the key ways that they are utilized: to improve the use of clustering as a model efficiency technique, allowing actuaries to improve their own efficiency and ability to support regulatory reporting. It also improves management decision-making on a flexible, timely, and cost-effective basis.

© 2022 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "MOODY'S"). All rights reserved.

CREDIT RATINGS ISSUED BY MOODY'S CREDIT RATINGS AFFILIATES ARE THEIR CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES, AND MATERIALS, PRODUCTS, SERVICES AND INFORMATION PUBLISHED BY MOODY'S (COLLECTIVELY, "PUBLICATIONS") MAY INCLUDE SUCH CURRENT OPINIONS. MOODY'S DEFINES CREDIT RISK AS THE RISK THAT AN ENTITY MAY NOT MEET ITS CONTRACTUAL FINANCIAL OBLIGATIONS AS THEY COME DUE AND ANY ESTIMATED FINANCIAL LOSS IN THE EVENT OF DEFAULT OR IMPAIRMENT. SEE APPLICABLE MOODY'S RATING SYMBOLS AND DEFINITIONS PUBLICATION FOR INFORMATION ON THE TYPES OF CONTRACTUAL FINANCIAL OBLIGATIONS ADDRESSED BY MOODY'S CREDIT RATINGS. CREDIT RATINGS DO NOT ADDRESS ANY OTHER RISK, INCLUDING BUT NOT LIMITED TO: LIQUIDITY RISK, MARKET VALUE RISK, OR PRICE VOLATILITY. CREDIT RATINGS, NON-CREDIT ASSESSMENTS ("ASSESSMENTS"), AND OTHER OPINIONS INCLUDED IN MOODY'S PUBLICATIONS ARE NOT STATEMENTS OF CURRENT OR HISTORICAL FACT. MOODY'S PUBLICATIONS MAY ALSO INCLUDE QUANTITATIVE MODEL-BASED ESTIMATES OF CREDIT RISK AND RELATED OPINIONS OR COMMENTARY PUBLISHED BY MOODY'S ANALYTICS, INC. AND/OR ITS AFFILIATES. MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND PUBLICATIONS DO NOT CONSTITUTE OR PROVIDE INVESTMENT OR FINANCIAL ADVICE, AND MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND PUBLICATIONS ARE NOT AND DO NOT PROVIDE RECOMMENDATIONS TO PURCHASE, SELL, OR HOLD PARTICULAR SECURITIES. MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND PUBLICATIONS DO NOT COMMENT ON THE SUITABILITY OF AN INVESTMENT FOR ANY PARTICULAR INVESTOR. MOODY'S ISSUES ITS CREDIT RATINGS, ASSESSMENTS AND OTHER OPINIONS AND PUBLISHES ITS PUBLICATIONS WITH THE EXPECTATION AND UNDERSTANDING THAT EACH INVESTOR WILL, WITH DUE CARE, MAKE ITS OWN STUDY AND EVALUATION OF EACH SECURITY THAT IS UNDER CONSIDERATION FOR PURCHASE, HOLDING, OR SALE.

MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS, AND PUBLICATIONS ARE NOT INTENDED FOR USE BY RETAIL INVESTORS AND IT WOULD BE RECKLESS AND INAPPROPRIATE FOR RETAIL INVESTORS TO USE MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS OR PUBLICATIONS WHEN MAKING AN INVESTMENT DECISION. IF IN DOUBT YOU SHOULD CONTACT YOUR FINANCIAL OR OTHER PROFESSIONAL ADVISER.

ALL INFORMATION CONTAINED HEREIN IS PROTECTED BY LAW, INCLUDING BUT NOT LIMITED TO, COPYRIGHT LAW, AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT.

MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND PUBLICATIONS ARE NOT INTENDED FOR USE BY ANY PERSON AS A BENCHMARK AS THAT TERM IS DEFINED FOR REGULATORY PURPOSES AND MUST NOT BE USED IN ANY WAY THAT COULD RESULT IN THEM BEING CONSIDERED A BENCHMARK.

All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, all information contained herein is provided "AS IS" without warranty of any kind. MOODY'S adopts all necessary measures so that the information it uses in assigning a credit rating is of sufficient quality and from sources MOODY'S considers to be reliable including, when appropriate, independent third-party sources. However, MOODY'S is not an auditor and cannot in every instance independently verify or validate information received in the rating process or in preparing its Publications.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability to any person or entity for any indirect, special, consequential, or incidental losses or damages whatsoever arising from or in connection with the information contained herein or the use of or inability to use any such information, even if MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers is advised in advance of the possibility of such losses or damages, including but not limited to: (a) any loss of present or prospective profits or (b) any loss or damage arising where the relevant financial instrument is not the subject of a particular credit rating assigned by MOODY'S.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability for any direct or compensatory losses or damages caused to any person or entity, including but not limited to by any negligence (but excluding fraud, willful misconduct or any other type of liability that, for the avoidance of doubt, by law cannot be excluded) on the part of, or any contingency within or beyond the control of, MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers, arising from or in connection with the information contained herein or the use of or inability to use any such information.

NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY CREDIT RATING, ASSESSMENT, OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.

Moody's Investors Service, Inc., a wholly-owned credit rating agency subsidiary of Moody's Corporation ("MCO"), hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by Moody's Investors Service, Inc. have, prior to assignment of any credit rating, agreed to pay to Moody's Investors Service, Inc. for credit ratings opinions and services rendered by it fees ranging from \$1,000 to approximately \$5,000,000. MCO and Moody's Investors Service also maintain policies and procedures to address the independence of Moody's Investors Service credit ratings and credit rating processes. Information regarding certain affiliations that may exist between directors of MCO and rated entities, and between entities who hold credit ratings from Moody's Investors Service and have also publicly reported to the SEC an ownership interest in MCO of more than 5%, is posted annually at www.moody.com under the heading "Investor Relations — Corporate Governance — Director and Shareholder Affiliation Policy."

Additional terms for Australia only: Any publication into Australia of this document is pursuant to the Australian Financial Services License of MOODY'S affiliate, Moody's Investors Service Pty Limited ABN 61 003 399 657AFSL 336969 and/or Moody's Analytics Australia Pty Ltd ABN 94 105 136 972 AFSL 383569 (as applicable). This document is intended to be provided only to "wholesale clients" within the meaning of section 761G of the Corporations Act 2001. By continuing to access this document from within Australia, you represent to MOODY'S that you are, or are accessing the document as a representative of, a "wholesale client" and that neither you nor the entity you represent will directly or indirectly disseminate this document or its contents to "retail clients" within the meaning of section 761G of the Corporations Act 2001. MOODY'S credit rating is an opinion as to the creditworthiness of a debt obligation of the issuer, not on the equity securities of the issuer or any form of security that is available to retail investors.

Additional terms for Japan only: Moody's Japan K.K. ("MJKK") is a wholly-owned credit rating agency subsidiary of Moody's Group Japan G.K., which is wholly-owned by Moody's Overseas Holdings Inc., a wholly-owned subsidiary of MCO. Moody's SF Japan K.K. ("MSFJ") is a wholly-owned credit rating agency subsidiary of MJKK. MSFJ is not a Nationally Recognized Statistical Rating Organization ("NRSRO"). Therefore, credit ratings assigned by MSFJ are Non-NRSRO Credit Ratings. Non-NRSRO Credit Ratings are assigned by an entity that is not a NRSRO and, consequently, the rated obligation will not qualify for certain types of treatment under U.S. laws. MJKK and MSFJ are credit rating agencies registered with the Japan Financial Services Agency and their registration numbers are FSA Commissioner (Ratings) No. 2 and 3 respectively.

MJKK or MSFJ (as applicable) hereby disclose that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by MJKK or MSFJ (as applicable) have, prior to assignment of any credit rating, agreed to pay to MJKK or MSFJ (as applicable) for credit ratings opinions and services rendered by it fees ranging from JPY100,000 to approximately JPY550,000,000.

MJKK and MSFJ also maintain policies and procedures to address Japanese regulatory requirements.