Compendium of the 23rd GCA papers

GCA24

DATA, DISRUPTIONS AND THE ACTUARY

12, 13, 14 FEBRUARY, 2024 AT THE WESTIN POWAI LAKE, MUMBAI

FOREWORD

Dear Readers,

With great enthusiasm, we are delighted to share the Compendium of the 23rd GCA papers. These papers will also be presented at the 23rd Global Conference of Actuaries on February 13, 2024. This initiative signifies our dedicated commitment to fostering academic research in actuarial science.

In our pursuit of excellence, we take immense pride in featuring four outstanding papers from the 23rd Global Conference of Actuaries in the compendium. These papers encapsulate cutting-edge insights, innovative methodologies, and thought-provoking analyses that reflect the depth and breadth of actuarial research presented at the conference.

We extend our gratitude to twenty peer reviewers from the Indian actuarial profession who peer-reviewed the entries and the sub-group comprising Ms. Rajeshwarie V.S., Mr. J V Prasad, and Mr. Abhijit Pal, whose collaborative efforts have significantly contributed to the selection and development of the shortlisted papers. The Institute of Actuaries of India looks forward to include these papers in an indexed journal in the short future.

This compendium will be easily accessible on our Institute's microsite, acting as a precursor to the full-fledged journal. We firmly believe in the power of knowledge dissemination, and this compendium marks our initial step towards bringing these insightful pieces to a wider audience.

The 23rd Global Conference of Actuaries carries the theme “Data, Disruptions, and the Actuary” to which the papers serve as a good fit.

Kulin Patel’s research provides useful insights on design options for the new pension scheme and raises the level of stimulating debate on the old versus the new pension scheme. Devadeep Gupta presents the new realm of actuarial science as it stands at the cusp of artificial intelligence, machine learning and broader data science techniques. Mitali Chatterjee underscores the importance and impact of digital trust in the era of data surfeit and AI whilst inferring that higher digital trust can translate into greater insurance density. KS Sandra Mary’s paper investigates the impact of a prospect’s profession on insurance choices, particularly in health insurance.

We trust you will find the compendium useful, and look forward to your feedback.

Warm regards,

R. Arunachalam

President, Institute of Actuaries of India

Anurag Rastogi

Chair, 23rd GCA Organizing Group

Mayur Ankolekar

Chair, Paper & Program Group, 23rd GCA Organizing Group

Mumbai. February 9, 2024.

INDEX

GCA24

DATA, DISRUPTIONS AND THE ACTUARY

12, 13, 14 FEBRUARY, 2024 AT THE WESTIN POWAI LAKE, MUMBAI

Paper I

Future of data science and machine learning applications in actuarial science

Future of data science and machine learning applications in actuarial science

Devadeep Gupta
FIAI CERA

ABSTRACT

Actuarial science is likely to evolve in the near future to embrace massive data volumes, uncertainties in future assumptions, and increasing complexity of processes and systems. Data science and machine can help provide effective tools to tackle these challenges.

Some examples where innovative techniques can help include:

- Use of machine learning to improve claims prediction and risk assessment.
- Anomaly detection techniques to enhance fraud detection.
- Synthetic data generation or data imputation to mitigate data availability issues.
- Natural language processing to help actuaries navigate system interfaces or complex processes.

The exponential growth in technology over the next five to ten years may amplify the need for these techniques to become not just a choice but a necessity for most companies. The paper will summarise how actuaries can design the overall strategy for a company to apply data science and machine learning.

KEYWORDS

Data Science, Artificial Intelligence, Machine Learning, Future of Actuarial Science, Insurance Technology, Fintech, Neural Network

CORRESPONDENCE DETAILS

Linkedin profile: https://www.linkedin.com/in/devadeepgupta/

Email: devadeepgupta313@gmail.com

INTRODUCTION

RELEVANCE OF DATA SCIENCE AND MACHINE LEARNING TO ACTUARIAL SCIENCE

In the past, the building blocks of actuarial science have been statistics, economics, risk management and an expertise in insurance, pensions, investment, banking or other financial domains. However, in recent decades the profession has developed significantly, with actuaries being required to familiarise themselves with applications in evolving actuarial modelling software and related technology platforms.

The world of finance is currently undergoing a profound transformation, fuelled by an exponential growth in data volumes and the rise of powerful **data science, artificial intelligence ("AI") and machine learning ("ML")** technologies.

In addition to this, several factors contribute to the increasing complexity of data required for business practices, including the following factors specific to insurance:

- The global insurance industry had reached US$6.3 trillion[1] in terms of total premiums in the year of 2019, with steady growth observed in both life and non-life premiums observed over past years. With the growth generating vast amount of data, this creates an unprecedented opportunity for actuaries to leverage data science and AI for deeper insights.
- The rise of connected devices and the Internet of Things ("IoT") has led to an explosion in the volume of data the insurance industry generates and stores.
- The insurance industry also faces pressure from competitors and consumers to use data more effectively and improve customer experience.

In line with this, this paper also explores some useful techniques in data science that are becoming increasingly relevant for the actuarial profession. Most of these techniques can also be applied outside of insurance, in banking, wealth management or pensions.

It is likely that some of the tools discussed in this paper are already in use in some global markets or multinational companies, while some techniques are still in the nascent stage of discussion and exploration. The aim of this paper is to help increase awareness of some of these techniques and suggest how common problems faced by the actuarial profession can be addressed using these techniques.

This paper aims to approach the actuarial profession from its roots as a science, by considering common real-life problems, and to expand our existing scientific techniques to use new tools under the bucket of data science and AI including machine learning.

This is but a natural evolution of the profession, as the application of some of the techniques described in this paper are likely to significantly impact and benefit core financial processes, such as product development and customer service.

POTENTIAL AREAS OF APPLICATION

Based on a survey by McKinsey, the number of AI capabilities that organisations use has doubled in the past few years – from an average 1.9 in 2018 to 3.8 in 2022[2]. More specifically within the field of insurance, 65% claims executives responding to an Accenture survey say they plan to invest more than $10 million into AI in the next three years[3].

There are quick wins and clear applications in cybersecurity or fraud prevention, which could lead to significant cost savings.

- **Security breach:** A study by IBM found that the average data breach cost for companies in 2022 was $4.35 million[4], with financial

[1]**Source:** Page 2 of Insurance Factbook 2021 https://www.iii.org/sites/default/files/docs/pdf/insurance_factbook_2021.pdf
[2]**Source:** https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review
[3]**Source:** https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-Why-AI-In-Insurance-Claims-And-Underwriting.pdf
[4]**Source:** IBM Corporation. Cost of a Data Breach Report 2022. Ibm.com, July 2022, https://www.ibm.com/downloads/cas/3R8N1DZJ

institutions averaging $5.97 million per breach. The study also found that it takes an average of 249 or 323 days - for companies with and without security AI and automation, respectively - for institutions to identify and contain a data breach.

- **Fraud detection:** In a 2022 study, the Coalition Against Insurance Fraud reported that insurance fraud could cost consumers $308.6 billion[5] annually in the United States alone. AI-based fraud detection systems could potentially reduce these losses.

However, beyond this, there are several other areas where data science and/or AI and ML may be applied by actuaries. In particular, this paper delves into applications which have the potential to revolutionise the profession. Some examples include:

- **Risk assessment:** Identifying risk patterns and optimising automated decision making
- **Pricing:** Supporting more accurate and personalised insurance premiums
- **Product innovation:** Designing new products tailored to specific needs and preferences of individual customers.
- **Customer support:** Supporting individual customers in giving them tailored financial planning advice.
- **Regulatory compliance:** AI can assist in complying with evolving regulatory requirements with automation and update of processes.

By embracing some of the techniques and expanding on the proposals in this paper, actuaries can enhance their predictive capabilities and optimise financial and strategic decision-making.

BACKGROUND AND SCOPE OF THE PAPER

This paper aims to approach the work of actuaries from a scientific point of view, focusing on some elements of actuarial tasks that are known to pose challenges today to key stakeholders. This will not be an exhaustive list of issues, or solutions, but the paper intends to increase awareness about these techniques so that actuaries get into the practice of assessing alternatives when faced with these situations, without getting intimidated by the novelty of some of these technologies.

While some of the issues discussed in this paper may have already been tackled through various alternative approaches in the past, the aim here is to objectively look at the problem from first principles and propose solutions using appropriate techniques.

In some cases, these techniques may already have been applied extensively in certain insurers or financial institutions, but it is likely that even in these cases, the same companies may look to the other solutions discussed in the paper as alternatives. In some cases, the proposed solution may remain theoretical at this stage, until a company or individual actuary chooses to apply it in their field of work.

The paper approaches the topic in the following structure:

- o **What are data science and machine learning? (Section: Introduction to data science and machine learning)**
- o **Why do we need data science and machine learning in actuarial science? (Section: Problem Statement)**
- o **How do we best use data science and machine learning? (Section: Potential Solutions)**
- o **What should actuaries and companies do next? (Section: Next Steps)**

As an illustration, the Problem Statement section identifies individual policyholders and customers of insurance companies and banks as a key stakeholder. Specific issues where data science and/or machine learning techniques may benefit them are discussed in the later sections. A similar approach is followed for other stakeholders, such as employees, senior management or shareholders of financial institutions.

Given the nature of the proposed techniques, the specific algorithm or model for any problem can often be applied to other problems in various sub-

[5]**Source:** Insurance Information Institute. Background on: Insurance Fraud. Iii.org, August 1 2022, https://www.iii.org/article/background-on-insurance-fraud

sections of this paper. This requires significant judgement to be applied by the actuary or data scientist going through the steps of the process. Hence, the potential solutions, case studies and practical next steps for companies should be treated as guidelines rather than rules to be followed for any particular issue.

INTRODUCTION TO DATA SCIENCE AND MACHINE LEARNING

This section aims to summarise the answer to the basic question "What are data science and machine learning?"

The key objective of this section is to clearly define the scope and boundaries of data science and machine learning, and in particular where the two concepts overlap in reference to potential applications in actuarial science.

Key terms are defined in this section, which are then referred to in following sections discussing the specific issues and solutions to practical problems where actuaries can apply these techniques. A detailed explanation of either AI or data science outside of the applications in actuarial science is out of scope.

Some techniques which have recently shown a significant increase in public interest, such as generative AI and deep learning, have specific sub-sections to explain where they sit in relation to the other terms and techniques discussed in this paper.

Data Science

Data Science refers to the use of scientific methods to obtain useful information from computer data, especially large amounts of data. It is an interdisciplinary field that uses statistics, computer science and a variety of algorithms to extract insights from potentially noisy, structured or unstructured data.

Data science involves the collection, cleaning, analysis and interpretation of data. The key activities within data science include:

- o **Data Collection:** Choosing data sources from websites, databases, sensors, etc and consolidating the data in suitable formats.
- o **Data Cleaning:** Checking and correcting the data to ensure it is accurate and consistent within acceptable thresholds. This may involve removing errors, filling in missing values and standardising formats.
- o **Data Analysis:** Using a variety of statistical and computational techniques to analyse the collected and cleaned data. This may involve exploring data, identifying patterns and building models.
- o **Data Interpretation:** Interpreting the results of analysis and communicating them in a clear way, with proposals on how to use the outcomes for business value.

Data science is used in a wide variety of industries and businesses, including finance and healthcare among others, to aid in better decision making about marketing, product development, risk management and treating diseases. Data science techniques are also used by some governments, to help in policy making, resource allocation and infrastructure development.

TYPES OF DATA SCIENCE TECHNIQUES

Data science may involve several types of techniques to perform the tasks mentioned earlier. Apart from machine learning algorithms, which we will delve into the next section, data science may be used in a variety of ways:

- **Predictive Analytics:** Data science can be used to perform predictive analytics by financial firms, for instance to assess customers' creditworthiness or likelihood to claim or surrender an insurance policy.
- **Natural Language Processing ("NLP"):** Large institutions or governments can use these techniques to analyse data from social media comments or public forums, to gauge public sentiment or target marketing campaigns.
- **Exploratory Data Analysis:** Summarising key characteristics from a large volume of data, and creating visual dashboards, including testing hypotheses and assumptions about the data. This can be done as the first step for any machine learning projects which will be described in later sections.

MACHINE LEARNING

Machine learning is a subfield of artificial

intelligence that focuses on enabling computer systems to learn and adapt without explicitly being programmed. It involves algorithms and statistical models to learn complex patterns from data and use it to draw inferences for decisions and predictions.

In contrast, AI which is not machine learning may rely on rule-based systems, where predefined rules and logic govern the system's behaviour.

While other AI methods may work where the rules can be fully defined in advance, machine learning allows systems to adapt and generalise to new and unseen data. This makes machine learning more versatile in handling complex pattern recognition and decision-making tasks.

Common machine learning application examples in everyday life include:

- o Image recognition algorithms on websites
- o Natural Language Processing used in customer service chatbots
- o Speech recognition in mobile phone applications
- o Recommendation systems on streaming platforms.

Types of machine learning – supervised, unsupervised and reinforcement

Supervised learning involves training a model on a labelled dataset, where the algorithm learns to map input data to corresponding output labels. This method is commonly used for tasks like classification and regression.

Unsupervised learning operates without labelled data and seeks to identify patterns or structures within the input data. Common applications of unsupervised learning include clustering and dimensionality reduction.

Common applications of both supervised and unsupervised learning in finance include credit scoring, fraud detection and algorithmic trading, but the two approaches work in different ways as noted above.

For example, dimensionality reduction (an unsupervised learning technique) can help simplify the analysis involved in algorithmic trading, without being overwhelmed by the diverse data involved in a diverse set of assets. On the other hand, algorithmic trading can also be performed by using classification algorithms (a supervised learning technique) to analyse market data and predict future asset price movements and automate buy and sell decisions based on price.

Reinforcement learning is a more dynamic and iterative approach where the system learns to make decisions by interacting with an environment. The system receives feedback in the form of rewards or penalties based on its actions, enabling it to optimise its decision-making strategy over time. This technique is often employed in applications such as game playing and autonomous driving.

In finance, taking the same example of algorithmic trading, reinforcement learning algorithms can learn trading strategies by interacting with simulated or real market data. Actions can be adjusted based on rewards (profits) or penalties (losses), gradually developing an effective strategy.

Reinforcement learning is a more like a trial-and-error approach, making independent decisions based on its understanding of the market and defined rewards and penalties.

Table 1: Summary of types of Machine Learning

Feature	Super-vised Learning	Unsuper-vised Learn-ing	Reinforce-ment Learn-ing
Data Type	Labelled Data	Unlabelled Data	Either type
Learn-ing Goal	Map input data to output data	Identify patterns and structure in data	Maximize rewards

RELATIONSHIP BETWEEN MACHINE LEARNING AND DATA SCIENCE

While machine learning is a specific approach within the broader field of AI, data science is a more comprehensive field that encompasses various techniques for handling and making sense of data.

The simplest way to relate the two would be to consider machine learning as one of the tools within the toolkit of data science. Data scientists, or actuaries and other professionals who use data science, may use machine learning algorithms, along with other methods, to extract valuable insights from data.

Example Projects

Table 2 summarises a few examples of how data science and machine learning may be applied in fields outside of actuarial science, simply to illustrate the definitions noted earlier in the section.

Table 2: Examples of data science and machine learning projects

Project	Data Science?	Machine Learning?	Project Objectives
Predictions for Equipment Failure	**Yes**	**Yes**	Use data science and machine learning algorithms to predict when equipment maintenance is required in factories to prevent failures
Exploratory Sales Trends Analysis	**Yes**	No	Analyse historical sales data manually to identify trends, patterns and insights without necessarily building a predictive model.
Recognition of digits from handwritten notes	No	**Yes**	Develop a machine learning model to accurately identify and classify handwritten digits (0 to 9) using image recognition techniques.

In summary, data science and machine learning are both broad fields with a variety of techniques, and can be applied together or separately, depending on business objectives.

Deep Dives

In this section, we look deeper into explaining some sub-topics within AI or data science. This is not an exhaustive summary of all potential techniques, but we look to cover some major topics that may be more relevant for insurance and banking solutions.

Neural Network

A neural network, or an **artificial neural network** to be more precise, is a type of machine learning algorithm inspired by the human brain. The ability of artificial neural networks to learn complex relationships between variables through their interconnected structure makes them particularly useful for modelling complex data analysis problems.

Neural networks consist of interconnected nodes called neurons, arranged in layers which process data and produce outputs. Neurons are connected by weights, and each connection has an associated weight that adjusts during the training process. **Neural networks usually have an input layer, one or more hidden layers where the inputs are processed, and an output layer. The model learns by adjusting the weights and connections between neurons based on training data.**

Neural networks can be applied in insurance and banking to solve a variety of problems.

As an example, a neural network for fraud detection may ingest transaction data (amount, location, date, etc), process the data and learn patterns within the hidden layers, and produce a binary result in the output layer (fraud or not fraud).

Another example would be in credit scoring, where the input layer may have customer and market data (credit score, financial history, economic data, etc), again analysing complex relationships within hidden layers, and producing a risk level or probability of default in the output layer.

There are various types of neural networks, some of which will be covered in practical examples in later sections. Neural networks may become powerful tools for actuaries in the future, especially if more financial companies embrace the technology and there is a demand for skilled employees who understand the principles behind the sophisticated internal mechanism.

Generative AI

Given the recent attention received by Generative

AI, a common question at this point may be "Where does Generative AI sit within the realm of AI and machine learning"?

Generative AI and machine learning are related concepts, but they refer to different aspects of AI. Table 3 summarises a comparison between the two. This paper focuses on the broader concept of machine learning (including potential applications of generative AI).

Table 3: Comparison of machine learning and generative AI

AI Type	Machine Learning ("ML")	Generative AI ("Gen AI")
Definition	A broader concept, including techniques and algorithms allowing computers to learn from data and make predictions or decisions without explicitly being programmed to do so.	A specific type of AI designed to generate new, original content or data. It creates models to produce data that is similar to, but not identical to, the training data provided.
Objective	Training models to perform specific tasks, such as regression or classification, based on patterns learned from data.	Generating new outputs, such as images, text or audio, that mimic the characteristics of training data.
Examples	Machine learning techniques can be used for: • Prediction of real estate prices • Classifying spam emails • Recognising images of cars.	Generative AI can be used for: • Generating realistic images of human faces • Creating new music, videos or text based on prompts.

Generative AI models can indeed leverage machine learning, especially in the training phase when they learn patterns from data. But their objectives are usually different. ML focuses on completing tasks, while Gen AI creates new content.

In summary, Gen AI and ML are different concepts but they can be applied together if the situation demands it. For example, both Gen AI and ML can come together to design a chat bot to generate human-quality text, such as those used for online customer support. Such a project may even involve training a recurrent neural network (a neural network designed to process a sequence of data) on a large corpus of text data.

As we will see in later sections, real-world problems faced by actuaries may lead us multiple potential techniques to choose from, and it is best not to think of the concepts as mutually exclusive.

DEEP LEARNING AND TRANSFER LEARNING

Deep Learning

Deep learning is a subset of machine learning, which is best used for complex tasks requiring machines to make sense of unstructured data. Deep learning typically uses neural networks with multiple layers, capable of automatically learning hierarchical representations of data.

This makes deep learning well-suited for tasks like image and speech recognition, and complex pattern recognition.

Transfer Learning

Transfer learning is a machine learning technique where a model trained on one task is adapted or fine-tuned for a different but related task. This saves the effort of training a model from scratch, leveraging knowledge gained from one domain and applying it to another to enhance performance.

This technique is particularly useful when there is limited data available for the final target task, but a similar model is available and pre-trained. An example would be where a company already has a pre-trained model to provide personalised advice to customers for individual investment strategies, but now they need a new model to advice customers on suitable insurance plans fitting their financial goals and risk appetite.

Both deep learning and transfer learning can offer valuable tools to actuaries in insurance and banking. While deep learning is perfect for situations involving large data volumes with complex relationships, transfer learning works best when there is an available model for a similar problem.

NATURAL LANGUAGE PROCESSING

Natural Language Processing ("NLP") is a subfield of AI that focuses on the interaction between humans using natural language to communicate with computers. NLP often involves the use of machine learning techniques to build models for tasks such as language translation, sentiment analysis and text summarisation.

NLP often involves generative AI techniques to generate human-like language, which is useful for communication with non-coders and users of systems such as customers. Large language models ("LLMs") such as GPT (Generative Pre-trained Transformer) models are a prominent example of generative AI used in NLP. These models are trained on vast amounts of text data.

NLP can be used in insurance and banking in personalised customer service through chatbots and virtual assistants, as well as in automating document review and flagging suspicious text data.

Choice of machine learning versus other AI in insurance and banking

While machine learning is only a subset of AI techniques, ML techniques are often the preferred choice over other techniques because machine learning models can learn from data and learn to make informed decisions. They can be trained to assess and predict risks, such as likelihood of customer default or a fraudulent claim. This gives ML an advantage over other AI methods, which are mostly rules-based systems and lack the adaptability and scalability or machine learning.

PROBLEM STATEMENT

Section Objective

This section explores the question "Why do actuaries need data science and machine learning?" A more tangible form in which we aim to answer this question is: "For whose benefit do we need data science and machine learning?"

The key stakeholders for the applications discussed in this paper generally fall within one of the following categories:

- Customers
- Investors
- Sales channels
- Company staff and management, including:
    - Senior executives
    - Actuaries
    - Data analysts
    - Accountants
    - IT team
    - Claims department
    - Underwriting department
    - Marketing department
    - Risk & Compliance department
- Reinsurers
- Competitors
- Supervisory authorities.

This paper will explore a few of these stakeholders in detail, but the list of challenges and solutions proposed are not meant to be an exhaustive list. The common challenges faced by each group will be considered and data science or machine learning techniques proposed to solve them. There will be a particular focus on the following areas:

- **Customers** of insurance companies and banks may benefit significantly in coming years if data science and ML techniques are applied to mitigate key risks.
- **Actuaries, data analysts and risk professionals** working across multiple departments within companies stand to gain from a knowledge a broader set of algorithms and techniques in assessing risk, pricing policies and automating processes, than is used currently.
- In particular, for tasks such as **underwriting, product design and claims handling,** ML can help study complex risk factors and detect frauds in a more efficient manner, supporting informed decision-making. Insight drawn from personalised and data-driven insurance or banking products are likely to benefit product design and development teams.
- **Sales and marketing teams** can benefit from more targeted campaigns, aimed at customer

preferences and concerns. Personalised recommendations are likely to add more value to customers of insurance companies and banks, and lead to better customer retention and persistency experience, while meeting customers' needs better.

- The **CEO and Board** of Directors may benefit from additional data-driven insights to inform strategic decision making about investments, resource allocation and overall business strategy. In more detail, the **CFO, Appointed Actuary and Chief Risk Officer** may be able to access improved financial forecasting and risk assessment, to aid them in decisions on capital allocation and risk mitigation strategies. These stakeholders are not discussed separately in this paper, as they are likely to benefit from all of the tools for the points mentioned above.

Customers

In this section, we will explore some of the key challenges faced by customers – a primary stakeholder for insurance companies, banks or other institutions where actuaries may apply data science and machine learning techniques.

The key challenges that customers may face, where the application of data science and/or machine learning may be relevant, include:

- o **Fraud:** This includes fraud related to transactions on credit cards, insurance claims, as well as broader risks such as identity theft.
- o **Security breach:** Customers are prone to data leaks, cyberattacks and scams such as phishing that pose a threat to loss of confidential and personal data.
- o **Lack of convenience and personalisation:** Customers may need extensive support in understanding their risks and the services or products that may be best suited for their specific situation. They are also likely to need support with product recommendations, basic admin support and proactive risk management.

Investors and Company Staff (including senior management & Actuarial Staff)

Departments within insurance companies and banks face a variety of issues in performing their day-to-day activities.

Actuaries working in Reporting, Pricing, Risk Management

Actuaries often face challenges in analysing the available data to produce insights in time for effective decision making. In particular, due to tight reporting timelines and competitive pressures, there is often a need for senior management including actuaries to see analytics real time. Monitoring key performance indicators (KPIs), and analysing them in time for investor communication, is often a challenging process in practice. There is a need to look at ways to optimise any potential automation in financial reporting and analyses processes, given they are a key input to strategic decision making. These processes also support risk management at corporation level, and hence are crucial for any insurance company or financial institution.

Some challenges faced by actuaries today during the actuarial processes (reporting, pricing, underwriting) include:

- Lack of granular data for analysis, often requiring manual effort to extract, analyse and test data
- Increasingly complex interfaces to systems and processes, introducing key man risk for processes requiring an in-depth understanding of underlying algorithms
- Need for real-time decision making for advice to senior management, or to departments such as underwriting or risk assessment, requiring ever-improving predictive models.
- Dynamic market conditions often make it necessary for models to adapt quickly with economic and risk factors, without significant time for manual effort and validation.
- Need for system validation tools for actuaries to analyse complex systems, such as design and development of independent models and checking techniques
- Lengthy process for generating and testing code from conceptual design or enhancement of financial models
- Need for extremely robust controls at multiple stages of the financial reporting process
- Need for robust cybersecurity and controls around confidential or personally identifiable data
- Challenges in pricing and underwriting

processes due to lack of granular risk assessment for potentially risky customers.

- Slow approval process in some cases, requiring extensive manual intervention.
- Fraudulent or malicious applications requiring sophisticated detection methods to avoid financial losses and reputational damage.

Claims Department

Apart from the above issues, there may be some specific challenges for the claims department, such as:

- Significant manual effort in processing claims, requiring manpower and resources, affecting efficiency.
- Fraudulent claims exacerbate the effort required, and at times it is challenging to identify such transactions, potentially leading to losses or unfair pay outs. Anti-fraud departments within companies now face increasingly sophisticated fraud schemes in the market, also leading to scalability issues with handling larger transaction volumes. Traditional fraud detection methods require regular upgrades to keep up with this growing challenge and defend against new tactics and detect evolving fraud patterns.
- Delays in resolution of claims due to slow approval or investigation processes have a knock-on effect on customer satisfaction and company reputation.

Sales and Marketing Department

Additional issues specific to the sales and marketing department may include:

- Inaccurate targeting of marketing campaigns, leading to higher costs than necessary and poor effectiveness of meeting customer needs.
- Exceedingly high customer acquisition costs, if sales process and initial on-boarding steps are expensive.
- Poor customer retention in case the evolving needs of individual policyholders are not catered to. In particular, customers often need extensive support as they may interact with the company through various channels. A seamless experience is the ideal target, which requires data analysis from multiple touchpoints to understand customer behaviours, preferences and issues.

In the next section, we shall look at how these challenges can be solved through data science and machine learning.

METHODOLOGY

Each company will have a multitude of issues for their actuaries to resolve, e.g. data issues, modelling issues, customers facing security issues or poor service.

While there is a significant amount of literature exploring specific techniques for a few of these issues, the key hypothesis of this paper is that there needs to be an overall Data Science or AI implementation strategy for each company, to avoid an exceedingly complex long-term IT infrastructure.

As multiple options exist in terms of data science or machine learning techniques to resolve the problems mentioned in the previous section, the first step followed was to categorise these issues into buckets, as noted in Figure 1.

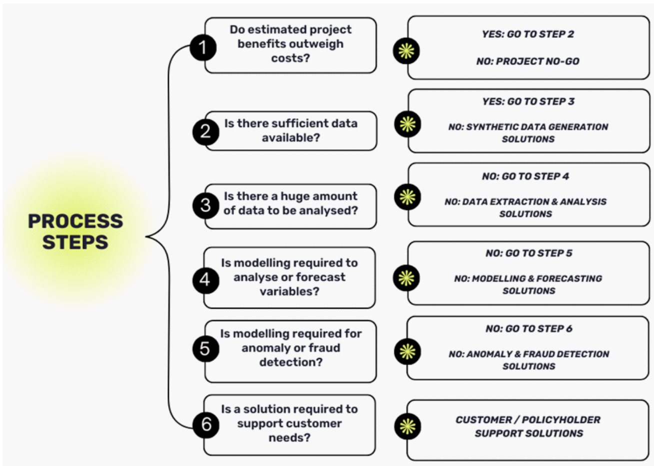


Figure 1: High level process steps to identify solutions for identified problems.

Once the categories have been formed, the various options for techniques that may be applied will have to be studied, in order to test performance of each option against each other and against existing models or processes.

As an example, if there is insufficient insurance claims data for a life actuary to perform mortality forecasting, they may end up studying alternatives and come up with the following list of “Synthetic Data Generation Solutions”. Please note that Figure 2 provides an example of what the solutions may look like for this particular issue. Case studies considered later this paper, combined with the list of techniques noted in the Potential Solutions section, then conclude with the full list of potential solutions mentioned in Figure 1.

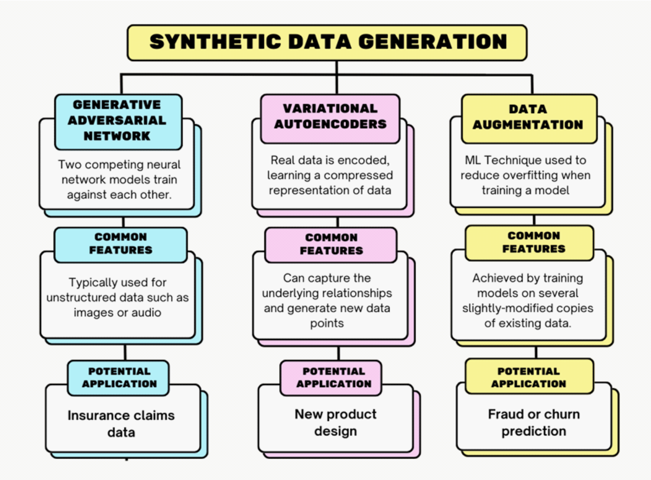


Figure 2: Sample list of Synthetic Data Generation Solutions

OVERALL AI STRATEGY

Before we apply the above process steps and come up with a more comprehensive list of solutions for all categories of problems identified earlier, a key question we look to tackle in this section is: "Why is there a need to have a single overall AI strategy?"

This is because having a siloed approach to data science and machine learning within an insurer or bank can lead to various risks and challenges, including:

- **Exceedingly complicated Technology Architecture:** Silos can lead to the development of independent and isolated systems for various departments or business units. This can lead to inefficiencies and even a risk to timely production during the BAU process in the long term.
- **Duplication of Effort:** Without collaboration and sharing of knowledge and resources, different teams may duplicate effort, leading to redundant processes, poor employee or customer satisfaction and increased costs.
- **Data inconsistency:** As data is a key piece of the puzzle, regardless of whether sophisticated models are used, it is necessary to consider minimum best practices, including consistency, confidentiality and accuracy of data used. This might prove difficult if, for example, some teams generate synthetic data not available to other teams.
- **Multiple black boxes:** Perhaps the biggest driver of long-term hidden costs and inefficiencies will be the risk of having multiple black boxes within the company infrastructure. As some of these techniques become more commonplace in banks and insurers, a strong core team to interpret and debug interpretability will be key for companies' success.

While a siloed approach may be required in some cases, with development of new models within a single team without involving many departments, information sharing and a centralised view should be ensured as far as possible. In particular, some customisation may be required for specific country regulations and to allow for domain expertise within a department e.g. underwriting.

Ultimately, the key is to strike a balance, depending on the company's structure, and the specific needs of different departments.

POTENTIAL SOLUTIONS

As the previous sections have summarised what we mean by data science and machine learning, and also identified the potential problems where these may be applied, **this section now discusses "How do we best use data science and machine learning to solve the identified problems?"**

In this section, we consider the challenges identified for various stakeholders in the previous section and explore potential solutions for each respective issue.

Customers

Fraud Detection

Various options exist to deal with anomaly detection, to help identify fraudulent transactions on credit cards, insurance claims and other cases, including:

- **One-Class Support Vector Machines ("SVMs"):** One-Class SVMs are a specific type of machine learning algorithm used for anomaly detection. While traditional SVMs are used to classify data into distinct categories, one-class SVMs focus on learning the characteristics of normal data to identify deviations or outliers. One-class SVMs are especially good at identifying fraudulent transactions, and even previously unseen fraudulent patterns. The technique is dependent on setting the right parameters and some review to prevent false positives.

- **Recurrent Neural Networks ("RNNs")[6]:** RNNs are a type of artificial neural network which, unlike traditional neural networks, can "remember" information from previous inputs in the sequence. RNNs can be used to model the temporal dependencies in credit card transaction data, and used to analyse spending patterns, flagging transactions deviating significantly from established baselines.
- **Gradient Boosting[7] :** This is a supervised learning ML algorithm, which combines the predictions of multiple weaker techniques such as decision trees, iteratively correcting errors and creating a strong predictive model. It can be trained on a labelled data set with instances marked as fraudulent or legitimate, learning patterns and ultimately predicting whether a new transaction is likely to be fraudulent.
- **Unsupervised learning algorithms[8] such as K-means clustering or Anomaly Detection Autoencoders:** These techniques can group data points based on their similarities (for K-means clustering) or compress and reconstruct data (for Anomaly Detection Autoencoders), potentially revealing patterns to signal anomalies or suspicious activity.

As there are a number of options available, actuaries or staff at insurance companies or banks should look to apply a combination of the best available methods for the specific case where they need to detect frauds. Attention must be paid to potential biases in the algorithm and errors in the outputs, so a continuous monitoring and improvement process will be necessary.

SECURITY BREACH

Security breaches can be dealt with in a number of ways to avoid loss of confidential and personal data:

- Intrusion detection systems powered by supervised learning models can analyse network traffic and systems logs and flag malicious or suspicious activity signatures.
- Unsupervised learning models like Isolation Forests or One-Class SVMs can detect anomalies in network traffic, system resource usage and user behaviour.
- Protection against phishing scams through Natural Language Processing ("NLP") combined with supervised learning models to analyse email content and website text to identify phishing attempts based on keywords, linguistic patterns and website reputation.

Convenience and personalisation

A number of options are possible to provide customer support in understanding their needs, risks and supply customers with tailor-made recommendations for their specific situation:

- NLP powered chatbots can analyse customer queries and intent, providing automated responses. Actuaries can work on the initial design of customer profiles and financial goals and risks, to improve the performance and effectiveness of chatbots. Continuous improvement is likely to be required.
- Machine learning models can be trained on dialogue datasets to improve chatbot responses and handle escalations to human agents.
- Collaborative filtering and content-based recommendation systems can analyse customer data (transactions, demographics, behaviour) to recommend appropriate products, services, or insurance schemes.

[6]Source: Page 12 of https://www.milliman.com/en/insight/the-use-of-artificial-intelligence-and-data-analytics-in-life-insurance
[7]Source: Page 40 of https://www.milliman.com/en/insight/the-use-of-artificial-intelligence-and-data-analytics-in-life-insurance
[8]Source: Page 12 of https://www.milliman.com/en/insight/the-use-of-artificial-intelligence-and-data-analytics-in-life-insurance

INVESTORS AND COMPANY STAFF (INCLUDING SENIOR MANAGEMENT & ACTUARIAL STAFF)

This section aims to identify solutions to the key challenges that company staff (including actuaries), as well as external shareholders, may face.

ACTUARIES WORKING IN REPORTING, PRICING, RISK MANAGEMENT

Data Challenges

While in some cases data availability can be offset by sourcing data for alternative sources, there may be occasions where significant manual effort is required to manufacture data manually, such as dummy data required for testing policies in a system prior to inception.

A potentially valuable solution would be synthetic data generation. This section details how such a case may be handled, with a set of proposed steps. Given this is potentially a common problem, we look at this problem in a bit more detail than other examples in this paper.

The scenario considered here is an **actuary facing a shortage of data in financial modelling of future liabilities for an insurer.**

Step 1: Identify data gaps

The first step is to list out and identify gaps in the data required for financial modelling. Just to illustrate how the machine learning solution may work, let us assume that the data gaps relate to future mortality rates, lapse rates and claims frequency and severity, in the situation where the actuary is trying to model the future liabilities for a life insurance company.

The required outputs required from this initial stage would be:

- o Clear quantification of required data volumes for modelling
- o Desired level of granularity for each data point, for e.g. mortality rate for each age split by sex, geography and other suitable measures

Step 2: Data Preparation

Once the required data and data gaps have been clearly quantified and documented, any existing data should be gathered, such as information from public sources, life tables, and economic or health trends available to the actuary.

Once the data is gathered, an important next step is to validate, analyse and clean or correct the data. This includes identifying gaps in the data, filling in gaps where it is possible to do so, and identifying the remaining areas which need help from data science or machine learning techniques.

Step 3: Choosing technique or model

There are various data science techniques to choose from, for such a problem. In reality, a considerable amount of time may be spent analysing multiple techniques and testing their relative results. Here, we take a closer look at a broad range of techniques known as **synthetic data generation**[9].

Synthetic data generation has the advantage of being adaptable for a wide range of data gaps, and it can be used to efficiently generate large datasets. Data properties can be tailored to specific needs, and data can be anonymised or obfuscated if required for confidentiality or privacy concerns.

However, when choosing this technique, caution is required for any bias in the underlying model. In addition, judgement is still required as the data produced may not be a replica of real data, and some complexity of real data may be missed out.

Synthetic data generation can be performed using a number of techniques, some of which are:

- **Generative Adversarial Networks ("GANs"):** Two competing neural network models (generator and discriminator) train against each other. The generator creates "fake" data mimicking any real customer data, while the discriminator tries to distinguish between the real and fake data. GANs are

[9]Source: Page 269 of https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2870308
GANs and VAEs are both advanced techniques, and they need a detailed cost and benefit analysis when applied in a real situation.

typically used for unstructured data such as images or audio, so additional work may be required to determine how to train the model to produce data for liability modelling, such as expected claims and surrender information for customers.

- **Variational Autoencoders ("VAEs"):** Real data is encoded, and by learning a compressed representation of this data, VAEs can capture the underlying relationships and generate new data points.

Step 4: Train the Model

- Once the model is selected, any pre-processed real data available should be fed into the model.
- There might be fine-tuning required to make sure the hyperparameters (variables that can be defined to control how the models are trained) are optimising the quality and realism of generated data.
- Model performance should be validated by continuously comparing synthetic data to real data distributions.

In summary, synthetic data generation has a number of benefits in financial modelling, especially in case of limited real data.

However, as with other techniques described in this paper, the setting up and training of the models will require technical expertise and data science skills, to avoid the models becoming complex black boxes.

Depending on the limitations of any available synthetic data generation techniques, alternatives, such as data augmentation, simpler generative models, or transfer learning may also be considered.

Modelling Issues

Machine learning can be used to develop automated testing tools, generate code quickly from a set of design specification documents, or even create independent validation models to check the accuracy and fairness of existing systems and actuarial models. Actuaries can help define the validation criteria and interpret results of these tests.

A few types of machine learning techniques that can help actuaries with modelling. Depending on the specific scenario, the options include the following:

- **Neural networks:** Deep learning techniques[10], such as neural networks, can be employed for complex pattern recognition tasks. As an example, a deep learning model can be used in pricing and reserving processes to model future mortality rates[11]. A deep learning model can also be trained on financial transactions for customers, including details on individual customer income, expense and risk appetite, to build a model to provide customised personal financial planning advice.
- **Natural Language Processing ("NLP"):** NLP algorithms can extract relevant information from text documents, such as a Word document with design specifications for a new actuarial model, and help convert it into code. While this will still require careful review and involvement from actuarial and IT teams, the process may be sped up significantly.
- **Support Vector Machines[12] :** Support vector machines are supervised machine learning models that can be effective to create models that can forecast future financial values, by using linear or non-linear regression to model relationships between variables.
- **Ensemble Learning[13] :** This technique can be used to improve overall accuracy and robustness, by combining various constituent machine learning algorithms. Random Forests are examples on ensemble learning, and can be used to model expected future insurance claims, for example, by analysing available data and predicting claim severity and

[10]Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3218082
[11]Source: Richman, Ronald (2021). Mind the gap – safely incorporating deep learning models into the actuarial toolkit. See References.
[12]Source: Page 12 of https://www.milliman.com/en/insight/the-use-of-artificial-intelligence-and-data-analytics-in-life-insurance
[13]Source: Page 12 of https://www.milliman.com/en/insight/the-use-of-artificial-intelligence-and-data-analytics-in-life-insurance

frequency.

- **Reinforcement Learning:** This technique can be applied for dynamic decision-making. In theory, a dynamic asset-liability modelling strategy could use such a modelling algorithm. Human review and feedback can also be provided to the model to help in optimising investment strategies and managing solvency.

The choice of the technique will depend on the requirements, data quality, domain expertise of the team and financial characteristics of the company. Multiple options exist for each problem, and actuaries may want to further explore the points mentioned above and consider the solutions across problems. The above list is not exhaustive and is intended to show that a wide range of solutions are possible. As with any other techniques, ethical considerations including data privacy and confidentiality must be adhered to and be carefully reviewed before any new processes are signed off.

Pricing and Underwriting

The following are good options to develop more accurate underwriting and pricing models. Machine learning models can be used to perform more accurate risk assessment, to support pricing and underwriting, such as:

- **Gradient Boosting Machines ("GBMs"):** These powerful ensemble models can potentially model complex relationships between risk factors and claims.
- **Neural networks:** Fed with data from wearable devices and telematics, sophisticated neural networks can be built to track driving behaviour for care insurance, as an example. This can help price insurance premiums more accurately and fairly.
- **Natural Language Processing ("NLP")** can automate parts of the underwriting process, such as document review and fraud detection. NLP algorithms can extract relevant information from unstructured documents, such as policy applications, medical records or underwriting notes. Actuaries should spend time to review the automation rules to ensure the systems are compliant with all relevant regulations.

CLAIMS DEPARTMENT

The following are good options to help claims teams:

- Machine learning models such as **Optical Character Recognition ("OCR")** can be used to automate parts of the claims handling process and reduce operational costs, such as data entry and document review. Actuaries can help design the automation workflow while ensuring that financial objectives are still met. Actuaries can also help train these models with historical data and develop strategies for investigating flagged claims.
- **Named Entity Recognition ("NER")** models are a type of Natural Language Processing technique that can identify and classify entities (names, dates, locations) within text documents.
- Fraud detection models (explored earlier in the sub-section on customers) are likely to benefit the claims department, automating flagging of suspicious claims for further investigation.
- Image recognition algorithms such as **convolutional neural networks** (a type of neural network that can analyse images to extract higher-level features) can analyse and classify images, such as photos of damaged property.

SALES AND MARKETING DEPARTMENT

Additional techniques that may help the sales and marketing teams:

- Machine learning models, such as **clustering algorithms,** can develop customer segmentation models to identify most profitable segments. This can be then used to develop marketing campaigns tailored to each segment.
- **Random forest** models can help develop lead scoring models to identify promising leads with a higher likelihood of purchasing specific insurance products. Actuaries can help set the scoring criteria in these models to prioritise marketing efforts.
- **Neural networks** can be developed to create prediction models to identify customers with

higher surrender risk. Actuaries can help design and implement these models to help with targeted retention programs.

Choice of Platform

While it is not a focus of this paper, it is worth discussing in brief a key question that is likely to come up: **which platform is best to use?**

The best option for each real case will require a cost benefit analysis, considering the programming expertise required, range of features available from various options, and how any new systems will interact with the existing platforms and system architecture of the company.

While there is no single platform that will suit every problem, the key options include[14] :

- **R:** An open-source programming language widely used for statistical computing and data analysis.
- **Python:** A programming language that is easy to learn and has a large number of inbuilt libraries and tools for data science and machine learning.
- Various other platforms such as KNIME, RapidMiner, Orange, and Jupyter Notebook that offer web-based interactive environments for data science and ML tools for exploring and analysing data.
- More advanced tools like Apache Spark, TensorFlow or PyTorch can run machine learning algorithms on large datasets.
- Firms such as IBM or SAS may offer their own platforms for data science and ML solutions for insurance and banking.

Ideally, any company looking to build new tools or introduce a new platform to their system architecture should perform a comprehensive option analysis as mentioned earlier. This should include a broader set of initial options such as build an in-house solution, purchase an out-of-the-box market product or choose to not develop models until a certain market event or agreed internal risk threshold is breached. A detailed look at these platforms is not considered in this paper.

FURTHER CONSIDERATIONS WHEN CHOOSING A SOLUTION

Overall, there are multiple options within the realm of data science and AI / machine learning when it comes to the common challenges faced by stakeholders of financial institutions. The aim of this paper is to show that it is possible for actuaries to apply their domain knowledge of a particular problem, and then assess the techniques available to conclude how to proceed.

Some additional thoughts regardless of which solution(s) are implemented by actuaries in various organisations are:

- Actuaries who are already familiar with some of the techniques discussed in this paper should nonetheless consider the options they have in each case and perform a pros and cons analysis to align with senior management on the costs and benefits of each exercise.
- Those who are not familiar with these techniques should look to add these skills to their repertoire as some of these algorithms are already in use today and are likely to become more commonplace in the future.
- Continuous monitoring and human intervention is likely to be equally, if not more, important than today, before the implementation of these techniques. This is to ensure that the chosen methods are accurate, fair and compliant with relevant regulations. It is unlikely that the sophistication of these tools will be sufficient to operate in dynamic regulatory environments without human input in the foreseeable future.

In summary, actuaries are likely to continue playing a central role in the insurance, pensions and financial industries where they operate today. As data science and machine learning become more prevalent, the quicker a large number of actuaries adopt these methods, the more valuable the profession will remain in changing times. An increasingly automated, but inherently complex, set of algorithms and processes, demands a group of strong actuaries who understand the underlying concepts and intention of the models,

[14]Source: Page 17 of https://www.actuaries.org.uk/system/files/field/document/Practical%20Application%20of%20Machine%20Learning%20within%20Actuarial%20Work%20Final%20%282%29_feb_2018.pdf

without resorting to excessive reliance on black box systems.

ETHICAL CONSIDERATIONS

As mentioned previously in the paper, there are several ethical issues to be considered when applying data science and machine learning techniques. Some key risks include:

- **Transparency:** Data science, and machine learning models, in particular, can be complex. It is advisable to employ individuals who are skilled in these techniques, rather than assuming the automation can replace human involvement. In fact, as inherent processes in companies become more sophisticated to deal with increasing volumes of data, it is more essential to have internal experts to avoid operating black boxes.
- **Model bias:** Outputs from machine learning models may be affected by underlying biases in their training data, and may produce misinformation or inappropriate output, such as indulging in discrimination. This should be supervised through human intervention as the consequences to company or institution reputation can be serious.
- **Data Privacy:** When creating synthetic data in particular, there should be strict measures in place to adhere to data confidentiality privacy policies and compliance requirements to uphold individuals' privacy.

In the initial stages of implementation of new techniques, actuaries and other personnel at corporations should review carefully whether any of the above risks may materialise. Continuous monitoring and supervision are recommended to ensure the systems continue adhering to any policies and regulations.

Large organisations and government bodies are already releasing their own set of guidelines or regulations on ethical considerations around the use of artificial intelligence, and this is likely to be an area where rules evolve rapidly in coming years.

CONCLUSION AND NEXT STEPS

Summary of Case Studies of Practical Implementation

There is empirical evidence of initial success in using some of the techniques described in the Potential Solutions section to address the aforementioned challenges in the financial industry. However, based on publicly available information, it is unclear to what extent these techniques have already been applied in insurance companies or banks worldwide.

Case Study Ref # (Appendix)	Problem	Technique	Methodology	Conclusions
Case Study 7	**Mortality modelling**	**Machine Learning (a combination of techniques)**	In the paper "Mortality Risk Modelling with Machine Learning" by Laurène Martin (2020)[15], a simplified insurance market was simulated: Two insurers using different pricing strategies, one based on machine learning and the other based on traditional regression models.	A first experiment highlighted that all other things being equal considering a Machine Learning model seems to allow to gain market shares and thus beat a contestant with regression methods. The insurer using a machine learning model manages to obtain a loss ratio close to 100% while the traditional one makes large losses.

[15]Source: https://www.institutdesactuaires.com/docs/mem/a1c834b2b50a8592bfc120173a0c1db2.pdf

Case Study 8	**Mortality forecasting**	**Recurrent Neural Networks**	In the book “Statistical Foundations of Actuarial Learning and its Applications” by Mario V. Wüthrich and Michael Merz, there is a discussion on mortality forecasting using recurrent neural networks[16]. As mortality data has a natural time-series structure, RNNs can be used to extrapolate the stochastic process in the Lee-Carter mortality model into the future.	Multiple studies have been done to successfully fit a Long short-term memory network (a type of RNN) to the stochastic process in the Lee-Carter mortality model. More generally, one can design a neural network architecture that directly processes the raw mortality data.
Case Study 9	**Pricing Model Comparison**	**GBM**	In the paper “A Comparison of Gradient Boosting Machines and Generalized Linear Models for Non-Life Insurance Pricing” by Alexander Eriksson, a comparison of model performance from a statistical perspective showed that the GBM provides a significant boost in prediction accuracy compared to the GLM.	From the perspective of creating pricing tariffs, the GBM has strong capability to rank risk and prevent adverse selection in a non-life insurance pricing setting, outperforming the GLM.
Case Study 10	**Underwriting**	**Machine Learning**	**RGA** has explored the application of AI and machine learning to facultative underwriting[17]. A quick win was the set of repetitive tasks in the workflow for a typical underwriter, such as a review of data, PDF images, tables and text files.	Machine learning could be used for voice, image, face, character recognition and translation algorithms that can be stacked like building blocks to support underwriters. The algorithms were particularly helpful in locating important information within thousands of pieces of documents and flagging incorrect manual inputs.

[16]Source: Page 403 of book with open access: https://link.springer.com/book/10.1007/978-3-031-12409-9
[17]Source: https://www.rgare.com/knowledge-center/article/wired-to-underwrite-artificial-intelligence-and-underwriting

Case Study 6	**Insurance Data Availability**	**GAN (Synthetic Data Generation)**	In the paper "Synthesizing Property & Casualty Ratemaking Datasets using Generative Adversarial Networks", the authors explore the use of a GAN for insurance claim data[18]. In this paper, the authors presented, implemented and compared three different methods to synthesize insurance data. All of the methods were based on generative adversarial networks.	All three methods have advantages and disadvantages and needed further work to adapt for specific insurance companies. One method synthesized the most realistic data, generating data which was very similar to the real data. GANs were considered a promising tool for synthesizing and protecting private and important data.
Case Study 1	**Fraud Detection**	**Recurrent Neural Network**	During training, the model accepted a total of 33,065 trainable parameters and used those parameters for obtaining prediction results[19]. The dataset consisted of 6,362,620 online transaction records during COVID-19 and each record was formulated as a collection of several attributes.	The model achieved an accuracy of 99.87%, F1-Score of 0.99 and MSE of 0.01. It was deemed to perform significantly well in detecting fraudulent transactions.
Case Study 2	**Fraud Detection**	**Recurrent Neural Network**	• The RNN model could handle both Card Present and Card Not Present scenarios, rather than needing multiple models. • RNN model saved time on feature engineering. • The RNN model could handle time-series data.	The model achieved an overall fraud detection rate of 35%, a substantial improvement over the 10% of existing models based on classical machine learning[20]. The RNN model also had potential to outperform the 40% achieved by internal Gradient Boosted Tree based models.W

[18]Source: https://hartman.byu.edu/docs/files/CoteHartmanMercierMeyersCummingsHarmon_SynthesizingDataGANs.pdf
[19]Source: https://www.researchgate.net/publication/347478467_Detection_of_Fraud_Transactions_Using_Recurrent_Neural_Network_during_COVID-19
[20]Source: https://www.crc.business-school.ed.ac.uk/sites/crc/files/2020-10/E29-Deep-Recurrent-Neural-Networks-Martini.pdf

Case Study 3	**Fraud Detection**	**K Means Clustering**	In the paper "Hybrid Methods for Credit Card Fraud Detection Using K-means Clustering with Hidden Markov Model and Multilayer Perceptron Algorithm", the authors modelled a fraud detection system that would attempt to maximally detect credit card fraud by generating clusters and analysing the clusters generated by the dataset for anomalies[21].	The performance of two hybrid approaches in terms of the detection accuracy was compared, and K-means clustering with the two approaches outperformed each other depending on the accompanying model. More extensive testing with much larger datasets was deemed required to validate these results.
Case Study 4	**Cybersecurity**	**Isolation Forest**	In the paper "An Isolation Forest Learning Based Outlier Detection Approach for Effectively Classifying Cyber Anomalies", the authors propose a model based on the Isolation Forest algorithm[22] to identify cyberattacks. The algorithm relies upon the characteristics of anomalies, i.e., being few and different, in order to detect anomalies.	Experimental results showed that the classification accuracy of cyber anomalies has been improved after the removal of outliers. More testing with large data sets with more security feature dimensions in IoT security services was proposed.
Case Study 5	**Personalised Customer Support**	**Natural Language Processing**	**Lemonade**, the well-known digital insurance provider, extensively uses NLP for its customer chat engine[23]. Lemonade's AI powered chatbot, AI Maya, handles everything from collecting information and personalising coverage to creating quotes and facilitating payments[24]. The AI chatbot Jim handles claims, investigating most cases with limited human involvement.	In 2019, the two bots handled 420,000 customer conversations[25], while AI Jim handled nearly 20,000 claims. AI Jim collected the information required, triaged the claims, handled emergencies, flagged fraud suspects and assigned escalations to the Lemonade team. AI Jim paid out claims worth nearly $2.5 million, with zero human involvement.

[21]Source: https://www.researchgate.net/publication/287358941_Hybrid_Methods_for_Credit_Card_Fraud_Detection_Using_K-means_Clustering_with_Hidden_Markov_Model_and_Multilayer_Perceptron_Algorithm
[22]Source: https://www.researchgate.net/publication/348380019_An_Isolation_Forest_Learning_Based_Outlier_Detection_Approach_for_Effectively_Classifying_Cyber_Anomalies
[23] Source: https://www.lemonade.com/blog/lemonade-review-2017/
[24] Source: https://www.lemonade.com/blog/two-years-transparency/
[25] Source: https://www.lemonade.com/blog/the-sixth-sense/

Case Study 11	**Claims Analysis**	**Convolutional neural networks**	In the paper "Convolutional Neural Networks for vehicle damage detection[26]" by R.E. van Ruitenbeek and S. Bhulai, the authors extended past research on claims analysis by using a significantly larger dataset by extending images available on the internet. They then evaluated the model against domain experts and assessed the performance in a production environment.	The paper showed that deep learning is able to accurately detect and classify vehicle damages, on a comparable level to human domain experts, and compiled bounding boxes in more detail. Damage detection seemed to benefit from models that focus on small objects and contextual information.
Case Study 12	**Sales and marketing – campaign targeting**	**Random forest**	In the paper "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis" by Weiwei Lin et al. (2017)[27], the authors proposed an ensemble random forest algorithm based on Apache Spark which can be used for classification of insurance data. They collected data from a life insurer to analyse potential customers using the proposed algorithm.	Experiment results showed that the ensemble random forest algorithm outperformed SVM and other classification algorithms in both performance and accuracy within the imbalanced data, and it was useful for improving the accuracy of product marketing compared to the traditional artificial approach. The ensemble random forest algorithm was more suitable in the insurance product recommendation or potential customer analysis than traditional strong classifiers like SVM and Logistic Regression.
Case Study 13	**Chain-ladder reserving**	**Neural networks**	The paper "Neural Networks Applied to Chain-Ladder Reserving" by Mario V. Wuthrich (2018)[28] describes how neural networks can be used for chain-ladder reserving and how these networks are calibrated to data using the gradient descent method. A chain-ladder model for claims reserving was extended to include individual claims feature information. This extension was done by using a neural network model for the factors.	The model to was applied to insurance data which consisted of close to 5 million individual claims histories. This extended model had the advantage that we could analyse reserves on individual claims, which allows us, for instance, to capture changing portfolio mixes.

[26] Source: https://www.sciencedirect.com/science/article/pii/S2666827022000433
[27] Source: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8007210
[28] Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2966126

CONCLUSIONS

It is likely that, given each company is likely to have its own set of challenges and pressing issues where a data science or machine learning solution is required, companies involved may choose to test the solutions for only their most urgent requirements first. New models are only likely to be implemented if there is a clear business case, and this is likely to only be true for a few problems at a given period of time.

However, the key assertion of this paper is that in the medium to long-term future, most of these techniques are likely to start getting adopted, given the large number of areas explored where multiple techniques may positively affect stakeholders.

The overall AI architecture of a single company might start to look exceedingly complex, if there is no long-term strategy for studying and implementing the available techniques. In line with this, the proposal is to have a centralised view of the overall list of problems where machine learning or data science may be applied, with documented results of multiple techniques against existing models, to continuously monitor and improve model performance.

The key focus of this paper was to consider some common challenges faced by actuaries and other stakeholders and suggest some options on how to resolve them using techniques within the realm of data science and machine learning.

Table 4 summarises the key challenges and proposed data science or machine learning techniques that are worth exploring further. This is not an exhaustive list of challenges or techniques at all, as one of the key observations in this paper is that given the large number of techniques available, **every specific situation may indeed have multiple, even dozens, of algorithms to choose from. Actuaries should consider Table 4 as a starting point for the overall AI strategy and architecture, and also review the next steps proposed later in this section.**

NEXT STEPS

Finally, **we will consider what companies or institutions, or individual actuaries or data scientists, should do next to implement data science and machine learning techniques for enhancing their business processes.** It is possible that some readers have already applied these techniques, so this section is written with two alternatives in mind: one in which a company is new to AI and data science, and another where they are already using some of the techniques described in this paper.

Scenario 1: Company is New to AI and Data Science

If an organisation is new to AI and data science, while the first steps might feel daunting, the potential long-term benefits, such as improved fraud detection, personalised customer advice and risk assessment, can outweigh the costs is the project is well planned and managed.

The next steps for such a company would be:

1. Gap Assessment for Needs

- Existing processes (such as claims, assumption setting, product design and other areas explored in this paper) should be studied, to identify quick wins and potential longer term improvement projects.
- Projects with the highest relative benefit versus cost should be prioritised. It is critical to set specific, measurable and realistic project goals.
- Management endorsement for the project budget and goals should be formalised, with regular updates and monitoring of progress. There should also be a clear long-term plan to integrate any new systems with existing long-term infrastructure and IT architecture of the company.
- There should be a strong core team with a combined skillset of project management and delivery with relevant domain experts (actuarial, accounting, data science, technology). While it may be challenging to hire AI or data science experts, an alternative is to train up existing experts in other areas, such as actuaries or technology / data analysts to also learn new skills in machine learning or data science.

2. Data and Tools

- Gather available data from various sources, such as internal systems, customer databases, and external websites. Existing

Table 4: Summary of challenges and proposed solution (*list of solution options is not an exhaustive list)

<table>
<tr><th>Stakeholder (For Whom?)</th><th>Issue (Why?)</th><th>Solution Options* (How?)</th></tr>
<tr><td rowspan="3">Customers</td><td>Fraud Detection</td><td>• One-Class Support Vector Machines
• Recurrent Neural Networks
• Gradient Boosting
• K-means clustering
• Anomaly Detection Autoencoders</td></tr>
<tr><td>Security Breach</td><td>• Intrusion detection systems powered by supervised learning models
• Unsupervised learning models like Isolation Forests or One-Class SVMs
• Natural Language Processing (“NLP”) combined with supervised learning models to analyse email content and website text</td></tr>
<tr><td>Convenience and personalisation</td><td>• NLP powered chatbots
• Collaborative filtering
• Machine learning models trained on dialogue datasets
• Content-based recommendation systems</td></tr>
<tr><td rowspan="5">Company employees and investors</td><td>Data Availability</td><td>Synthetic data generation techniques, such as:
• Generative Adversarial Network
• Variational Autoencoders

Depending on the limitations of any available synthetic data generation techniques, alternatives, such as data augmentation, simpler generative models, or transfer learning may also be considered.</td></tr>
<tr><td>Modelling Issues</td><td>• Neural Networks, in particular deep learning techniques
• Natural Language Processing
• Support Vector Machines
• Ensemble Learning
• Reinforcement Learning</td></tr>
<tr><td>Pricing & Underwriting</td><td>• Gradient Boosting Machines
• Neural networks
• Natural Language Processing</td></tr>
<tr><td>Claims</td><td>• Machine learning models such as Optical Character Recognition
• Named Entity Recognition models
• Image recognition algorithms such as convolutional neural networks</td></tr>
<tr><td>Sales & Marketing</td><td>• Machine learning models, such as clustering algorithms
• Random forest models
• Neural networks</td></tr>
</table>

data gathering processes should be revisited, in case there is a need for extra effort to clean the data to train new ML models, or to constantly review the output and provide feedback in a reinforcement learning model.

- Table 4 can be a reference for a starting position for new models to be considered, but each company needs to assess their own needs and the performance of available models suited to their data and situation.
- An extensive amount of time should be spent on choosing the right tools or technologies (including platform choice, model choice and data infrastructure such as data warehouse, data lakes, etc). For a company which is new to AI algorithms, the initial choice may lead to investing in technologies that will remain within large organisations for a significant amount of time, so the initial options evaluation should be as comprehensive as possible, involving all key departments and affected stakeholders.

3. Continuous development in people, processes and systems

- Once any new models are designed and developed, constant monitoring and improvement will be necessary.
- This can be taken as a good opportunity to train existing staff and upskill them with real experience of working with machine learning models.
- As set out in the initial project plan, once any new systems or processes are in place, they should be seamlessly integrated with long-term BAU processes, including extensive system integration testing, followed by user acceptance testing. It is essential that the end users of the systems are also trained to use them to avail the originally intended benefits and continue to perform their daily activities through training on the new interfaces and algorithms.

By following these steps, a bank or insurer can integrate data science and machine learning into their operations. **The key may be to start with smaller or more realistic projects if possible, targeting known challenges in the existing processes, and then scale up following any findings and observations from the smaller pilot project.**

SCENARIO 2: EXISTING FRAMEWORK FOR AI AND DATA SCIENCE

For companies that already have some experience or existing processes using the techniques discussed in this paper, the process is broadly the same, with the following key differences:

- The focus shifts to enhancing and optimising the existing data science and ML practices, rather than starting from scratch.
- Existing models should be reviewed for their performance, with clear thresholds for errors and biases. Retraining or updates should be applied to maintain fairness to customers and shareholders. Iterative or agile approaches are likely to be necessary, with clear and precise communication to senior management on any issues and corrections required.
- Trainings should continue, as machine learning techniques become more widespread within each company, to raise base level awareness across all staff, as there may be a large number of users of the output from these systems. At the same time, care should be taken to retain and develop a core team of talent who are skilled at data science and ML techniques across the domains of Actuarial, Finance and IT teams. Building cross-functional teams who do not just stick to their own silos will become more necessary in the coming future.
- Use cases should be expanded beyond existing applications, wherever data science or ML can add more value. For example, if a company is only using AI for customer support chatbots but nowhere else, it is worth exploring if the techniques to help claims handling or product design discussed in this paper are worth exploring as well.
- Staff in relevant departments should be trained to continue to automate routine tasks and spend more time on analysis and gathering insights for management. This will lead to long-term cost savings and efficiency gains.
- Knowledge sharing on these techniques should be encouraged across departments. Collaboration with universities, research

institutions or software providers might become necessary in the medium to long term to avoid the processes falling behind most market players.

- Data governance and cybersecurity frameworks should be further strengthened. Clear guidelines should be in place for data access, storage and usage; and staff should be made aware of cyber threats and trained to guard against them in their daily job.
- Ethical guidelines should be established to support the development of any new data science or AI algorithm, given the risks noted earlier in this paper.
- Training top talent and retaining them may become more necessary, as in the short to medium term the demand for actuaries who are also familiar with ML techniques or data science might be in high demand. It is necessary to train up most staff in Finance and Actuarial on the basics of these techniques and encourage them to come up with proposals for new solutions to help the organisation.

By following these steps, companies are likely to keep up with and perhaps stay ahead of market trends in terms of the usage of these techniques in finance and insurance. As the techniques themselves evolve with the underlying risks, such as ever-more complex systems, larger data volumes and advanced security threats, continuous self-learning should be promoted by all actuaries so that the use of data science and machine learning becomes a regular part of our profession.

APPENDIX A: CASE STUDIES

Case Study 1: Fraud detection using Recurrent Neural Network ("RNN")

In the article Detection of Fraud Transactions Using Recurrent Neural Network during COVID-19 by Dutta S, Bandyopadhyay SK[29], the authors highlight the performance of a Recurrent Neural Network applied on Paysim generated synthetic financial dataset. The model detected deceptive transactions with an **accuracy of 99.87%, F1-Score of 0.99 and MSE of 0.01.**

Details of the model are as follows:

- A stacked RNN model was proposed as a recommender system for detection of fraud transaction.
- Multiple RNN layers were stacked into a single platform for obtaining the proposed model. Four simple RNN layers along with four dropout layers are incorporated into a sequential model.
- These layers were compiled using 'Adam'16 optimizer (Adam optimization is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments) and binary cross entropy loss function.
- During training, the model accepted a total of 33,065 trainable parameters and used those parameters for obtaining prediction results.
- The original transaction data was collected from a month's worth financial logs of a mobile money service implemented in an African country.
- The dataset consisted of 6,362,620 online transaction records during COVID-19 and each record is formulated as a collection of several attributes.

Case Study 2: Fraud Detection using Recurrent Neural Network ("RNN")

In the use case described in "Deep Recurrent Neural Networks for Fraud Detection on Debit Card Transactions" by Antonio Martini, the author discusses the application of an RNN for fraud detection, achieving an **overall fraud detection**

[29] Source: https://www.researchgate.net/publication/347478467_Detection_of_Fraud_Transactions_Using_Recurrent_Neural_Network_during_COVID-19

rate of 35%, a substantial improvement over the 10% of existing models based on classical machine learning[30]. The improvements were attributed to recent improvements in deep learning and neural network technologies. The RNN model also had potential to outperform the 40% achieved by internal Gradient Boosted Tree based models.

Some properties of the model were:

- The RNN model could handle both Card Present and Card Not Present scenarios, rather than needing multiple models.
- Features were derived using aggregated statistics (e.g. mean, count, sum over a time period) to provide temporal information, and this was a time consuming process for existing models but the RNN model saved time on this part of the process.
- The RNN model could handle time-series data, and the automatically computed RNN state efficiently summarised the time-series past. This handled a complex algorithm of allowing learning over long sequences as is the property of RNNs, allowing for relevant data to be stored in the long term memory.
- Training the model consisted of finding the RNN parameters that minimise the chosen cost function: cross-entropy, which is a measure of the performance of the classification performed by the model, by defining the difference between the estimated probability distribution and our desire outcome.
- The model was built and trained using TensorFlow, an open-source ML platform.
- The results were measured through the result of TDR (Transaction Detection Rate) for ALL, CNP and CP is evaluated at different operating points, which was calculated as: (Number of detected frauds within operating point / Number of frauds in data).

Case Study 3: Fraud detection using K means clustering

In the paper “Hybrid Methods for Credit Card Fraud Detection Using K-means Clustering with Hidden Markov Model and Multilayer Perceptron Algorithm”, the authors modelled a fraud detection system that would attempt to maximally detect credit card fraud by generating clusters and analysing the clusters generated by the dataset for anomalies[31]. The performance of two hybrid approaches in terms of the detection accuracy was compared.

The authors employed hybrid methods using the K-means Clustering algorithm with Multilayer Perceptron (MLP) and the Hidden Markov Model (HMM) for this study. The tests revealed that the detection accuracy of “MLP with K-means Clustering” is higher than the “HMM with K-means Clustering” for 80% percentage split but the reverse is the case when the “MLP with K-means Clustering” is compared with the “HMM with K-means Clustering” for 10-fold cross-validation but the accuracy is the same in the two hybrid methods for percentage split of 66%. More extensive testing with much larger datasets was however deemed necessary to validate these results.

Case Study 4: Cybersecurity using a variety of ML techniques

There is evidence of the potential application of Isolation Forests for anomaly detection, in particular to combat one of the largest emerging threats to both customers and financial companies: cybersecurity. In a study, an Isolation Forest was shown to detect anomalies using binary trees[32]. Isolation Forest is an unsupervised learning ML algorithm, using the average prediction by several decision trees when assigning an anomaly score to a given data point.

Supervised classification techniques can learn well from a security dataset that can play a major role in tracking network traffic and

[30] Source: https://www.crc.business-school.ed.ac.uk/sites/crc/files/2020-10/E29-Deep-Recurrent-Neural-Networks-Martini.pdf

[31] Source: https://www.researchgate.net/publication/287358941_Hybrid_Methods_for_Credit_Card_Fraud_Detection_Using_K-means_Clustering_with_Hidden_Markov_Model_and_Multilayer_Perceptron_Algorithm

[32] Source: https://www.researchgate.net/publication/348380019_An_Isolation_Forest_Learning_Based_Outlier_Detection_Approach_for_Effectively_Classifying_Cyber_Anomalies

taking corrective action when unauthorised transmissions are discovered. However, modelling cyber-attacks effectively is problematic due to the high dimensions of security features and the presence of outliers in today's security datasets, in line with the use of Isolation Forests which are unsupervised.

The algorithm typically has a linear time complexity and a low memory requirement, which works well with high-volume data. In essence, the algorithm relies upon the characteristics of anomalies, i.e., being few and different, in order to detect anomalies.

In the paper "An Isolation Forest Learning Based Outlier Detection Approach for Effectively Classifying Cyber Anomalies", the authors propose a model based on the Isolation Forest algorithm. The model involves the following steps:

- Performing necessary pre-processing steps like categorical feature encoding, feature scaling to extract fifteen essential features to fit into the proposed model. Firstly, the pre-processing is performed on raw data to normalize the values of the features.
- The dataset was created by simulating a standard local area network of the US Air Force, which is exposed to numerous cyber-attacks known as intrusions.
- The most common outlier detection techniques are focused on building a profile of "normal" instances, after which the outliers are recorded as those that do not adhere to the normal profile in the dataset.
- After detecting outliers using the Isolation Forest algorithm, all outliers were removed to evaluate the system's improved performance in terms of classification accuracy.

Case Study 5: Personalised Customer Support using Natural Language Processing ("NLP")

There is a wide variety of companies using Natural Language Processing to improve customer service. NLP is the core technology behind well-known virtual assistants such as Siri, Cortana or Oracle Digital Assistant[33].

Lemonade, the well-known digital insurance provider, extensively uses NLP for its customer chat engine[34]. Lemonade's AI powered chatbot, AI Maya, handles everything from collecting information and personalising coverage to creating quotes and facilitating payments[35]. The AI chatbot Jim handles claims, investigating most cases with limited human involvement.

In 2019, the two bots handled 420,000 customer conversations[36], while AI Jim handled nearly 20,000 claims. AI Jim collected the information required, triaged the claims, handled emergencies, flagged fraud suspects and assigned escalations to the Lemonade team. AI Jim paid out claims worth nearly $2.5 million, with zero human involvement.

NLP is likely to be one of the machine learning algorithms that is used more broadly in the future, with potential applications in providing customised financial advice to customers.

Case Study 6: Synthetic data generation with Generative Adversarial Network ("GAN")

In the paper "Synthesizing Property & Casualty Ratemaking Datasets using Generative Adversarial Networks", the authors explore the use of a GAN for insurance claim data[37]. The authors explored the application of improved computing resources in modelling individual claims and loss reserving models. A key hurdle in the training of new models is the confidentiality of sensitive customer information, making it difficult to compare methods to each other or to new methods yet to be developed.

In this paper, the authors presented, implemented and compared three different methods to synthesize insurance data. All of the methods were based on generative adversarial networks. All three methods have advantages and disadvantages and needed further work to adapt

33 Source: https://www.oracle.com/artificial-intelligence/what-is-natural-language-processing/
34 Source: https://www.lemonade.com/blog/lemonade-review-2017/
35 Source: https://www.lemonade.com/blog/two-years-transparency/
36 Source: https://www.lemonade.com/blog/the-sixth-sense/
37 Source: https://hartman.byu.edu/docs/files/CoteHartmanMercierMeyersCummingsHarmon_38SynthesizingDataGANs.pdf

for specific insurance companies. One method synthesized the most realistic data, generating data which was very similar (accounting for both univariate and multivariate relationships) to the real data. Future work could start from any of the three models and attempt to add the advantages of the other two.

In addition to this, various studies have been performed in the past for synthetic data generation or augmentation to improve classification and analysis of health records. Examples include:

- The paper "Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification" by Frid-Adar et al. (2018), exploring the use of GANs to generate synthetic data in order to augment a small imaging dataset and improve the performance of the classification of liver lesion[38].
- The paper "Generating Multi-label Discrete Patient Records using Generative Adversarial Networks" by Choi et al. (2017)[39], proposing an architecture to synthesize realistic patient records for developing new models and statistical methods.

The structure of patient record data is also closer to that of insurance data than most of the deep learning literature, focusing on unstructured data such as images. Images (and pixels) are continuous, whereas most of claimant characteristics are categorical variables, which adds complexity as one cannot interpolate between discrete classes to create fake records.

Case Study 7: Mortality Modelling for Pricing or Reserving using machine learning

In the paper "Mortality Risk Modelling with Machine Learning" by Laurène Martin (2020)[40], the author evaluates the benefits of mortality modelling with machine learning.

All the methods were implemented within a python library, and the library aimed at standardising several methods to facilitate and automate the study of the mortality of a portfolio.

To analyse the impact of different mortality modelling approaches, an open-source database was used, which was a program of studies originally designed to assess the health and nutritional status of adults and children in the United States. The dataset was composed of 65, 018 individuals, and 106 variables, which were categorized into five classes: demography, dietary, laboratory, examination, and questionnaire.

The library was developed in a way to standardise every model, which enabled and facilitated their comparison, as model performance was heavily dependent on available data sets.

A simplified insurance market was simulated, with two insurers using different pricing strategies for the same product, one using machine learning while the other used a traditional model. A first experiment highlighted that all other things being equal considering a Machine Learning model could potentially allow to gain market shares and thus beat a contestant model with regression methods. In the study, the insurer using a machine learning model managed to obtain a loss ratio close to 100% while the traditional one made large losses.

The division of the market between both insurers was beneficial to the one with the most advanced technology as this insurer was able to offer more attractive prices for less risky individuals.

While for these neural network models, there is a potential for them to be seen as "black box" Machine Learning models, which cannot be directly interpretable. However, in many cases, understanding the model remains essential for respecting the regulatory requirements, to justify business decision-making and keeping stakeholder trust by understanding the results produced, etc. For that purpose, multiple methods of interpretation are likely to be necessary.

It is likely to become more common for companies to study multiple options for each problem. No clear conclusion can be drawn on the model which will be preferred, until the options are tested on large volumes of data.

[38] Source: https://arxiv.org/pdf/1801.02385.pdf
[39] Source: https://proceedings.mlr.press/v68/choi17a.html
[40] Source: https://www.institutdesactuaires.com/docs/mem/a1c834b2b50a8592bfc120173a0c1db2.pdf

Case Study 8: Mortality Forecasting using Recurrent Neural Networks ("RNNs")

In the book "Statistical Foundations of Actuarial Learning and its Applications" by Mario V. Wüthrich and Michael Merz, there is a discussion on mortality forecasting using recurrent neural networks[41].

As mortality data has a natural time-series structure, RNNs can be used to extrapolate the stochastic process in the Lee-Carter mortality model into the future. Multiple studies have been done to fit a Long short-term memory network (a type of RNN) to this stochastic process. More generally, one can design a neural network architecture that directly processes the raw mortality data.

Case Study 9: Pricing Model using Gradient Boosting Machines ("GBM")

In the paper "A Comparison of Gradient Boosting Machines and Generalized Linear Models for Non-Life Insurance Pricing" by Alexander Eriksson, a comparison of model performance from a statistical perspective showed that the GBM provides a significant boost in prediction accuracy compared to the GLM.

From the perspective of creating pricing tariffs, the GBM has strong capability to rank risk and prevent adverse selection in a non-life insurance pricing setting, outperforming the GLM.

By using a measure of variable importance, we could gain insights into what variables the GBM regards as most important for prediction of the claim frequency, similar to the parameter weights of the Poisson GLM. Further, by application of partial dependence plots, there can be insights into how the variables in the GBM model affect the predicted claim frequency.

Case Study 10: Underwriting using Machine Learning

RGA has explored the application of AI and machine learning to facultative underwriting[42]. Machine learning could be used for voice, image, face, character recognition and translation algorithms that can be stacked like building blocks to support underwriters.

A quick win was the set of repetitive tasks in the workflow for a typical underwriter, such as a review of data, PDF images, tables and text files.

The time-consuming manual review of PDF images, requiring both visual review and transcribing work, could be addressed by optical character recognition or OCR, a form of artificial intelligence that can find patterns in images. Natural language processing could mine text for sequences of words. The algorithms were particularly helpful in locating important information within thousands of pieces of documents, and flagging incorrect manual inputs.

Case Study 11: Claims analysis using convolutional neural networks

In the paper "Convolutional Neural Networks for vehicle damage detection[43]" by R.E. van Ruitenbeek and S. Bhulai, the authors extended the past research on claims analysis by using a significantly larger dataset by extending images available on the internet. They also applied the damage classification on 12 categories of damage. They then evaluated the model against domain experts and assessed the performance in a production environment.

The paper showed that deep learning is able to accurately detect and classify vehicle damages, on a comparable level to human domain experts, and compiled bounding boxes in more detail. Damage detection seemed to benefit from models that focus on small objects and contextual information. Domain experts made slightly fewer false negatives for Dents, where the model outperforms in the classes Bend and Cover Damage. Additionally, the model is less accurate in identifying the background (no damage) from damage within the light street but is still able to distinguish the different damage classes.

[41] Source: Page 403 of book with open access: https://link.springer.com/book/10.1007/978-3-031-12409-9
[42] Source: https://www.rgare.com/knowledge-center/article/wired-to-underwrite-artificial-intelligence-and-underwriting
[43] Source: https://www.sciencedirect.com/science/article/pii/S2666827022000433

Case Study 12: Sales and marketing using random forest (campaign targeting)

In the paper "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis" by Weiwei Lin et al. (2017)[44], the authors proposed an ensemble random forest algorithm based on Apache Spark which can be used for classification of insurance data.

It is at times difficult to model the insurance business data by classification algorithms, such as logistic regression and support vector machine (SVM), and the paper looked to exploit a heuristic bootstrap sampling approach combined with the ensemble learning algorithm on the large-scale insurance business data. They collected data from a large life insurer to analyse potential customers using the proposed algorithm.

Experiment results showed that the ensemble random forest algorithm outperformed SVM and other classification algorithms in both performance and accuracy within the imbalanced data, and it was useful for improving the accuracy of product marketing compared to the traditional artificial approach. The ensemble random forest algorithm was more suitable in the insurance product recommendation or potential customer analysis than traditional strong classifiers like SVM and Logistic Regression.

Case Study 13: Chain-ladder reserving using neural networks

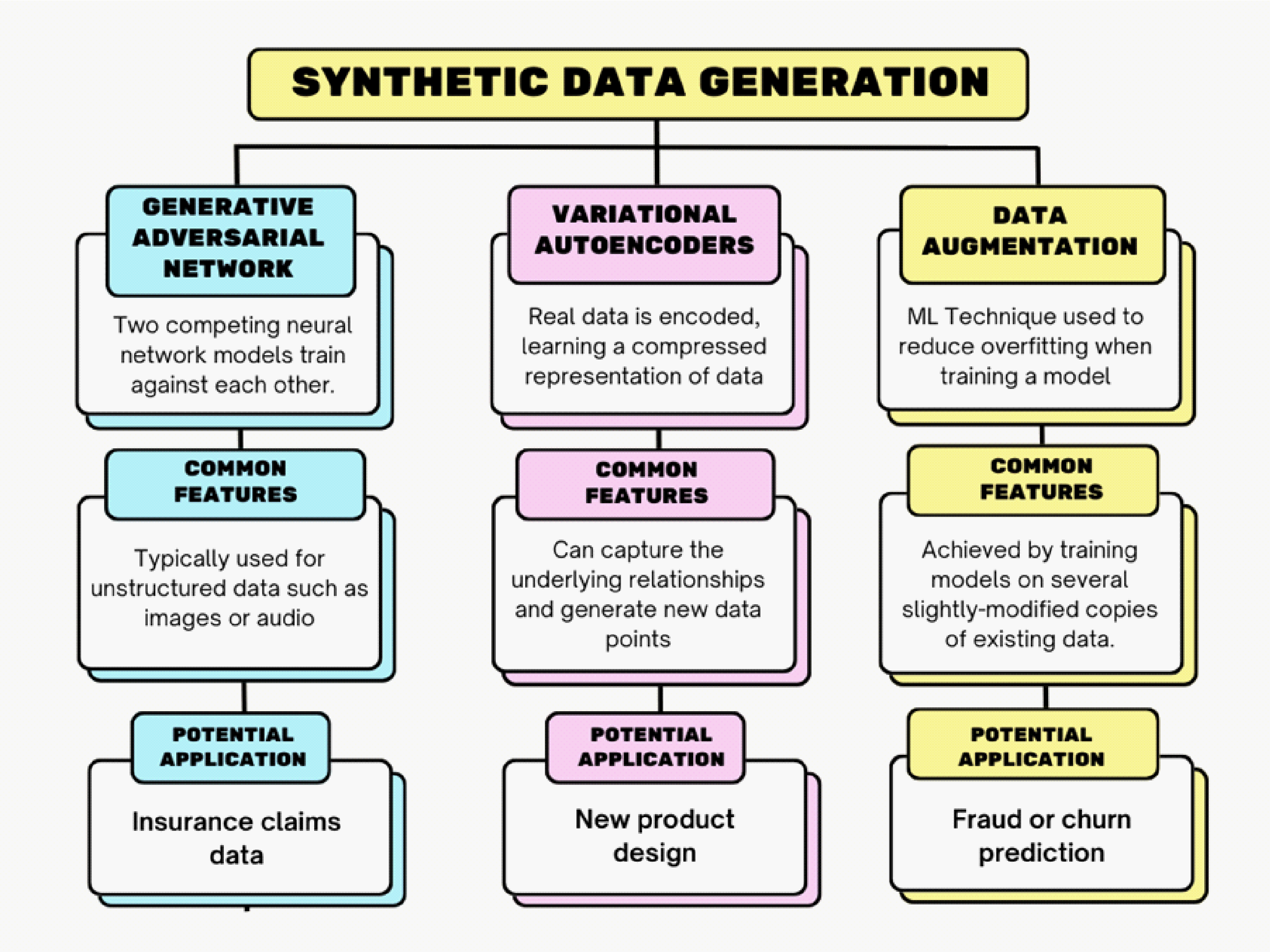
The paper "Neural Networks Applied to Chain-Ladder Reserving" by Mario V. Wuthrich (2018)[45] describes how neural networks can be used for chain-ladder reserving and how these networks are calibrated to data using the gradient descent method. The model to was applied to insurance data which consisted of close to 5 million individual claims histories.

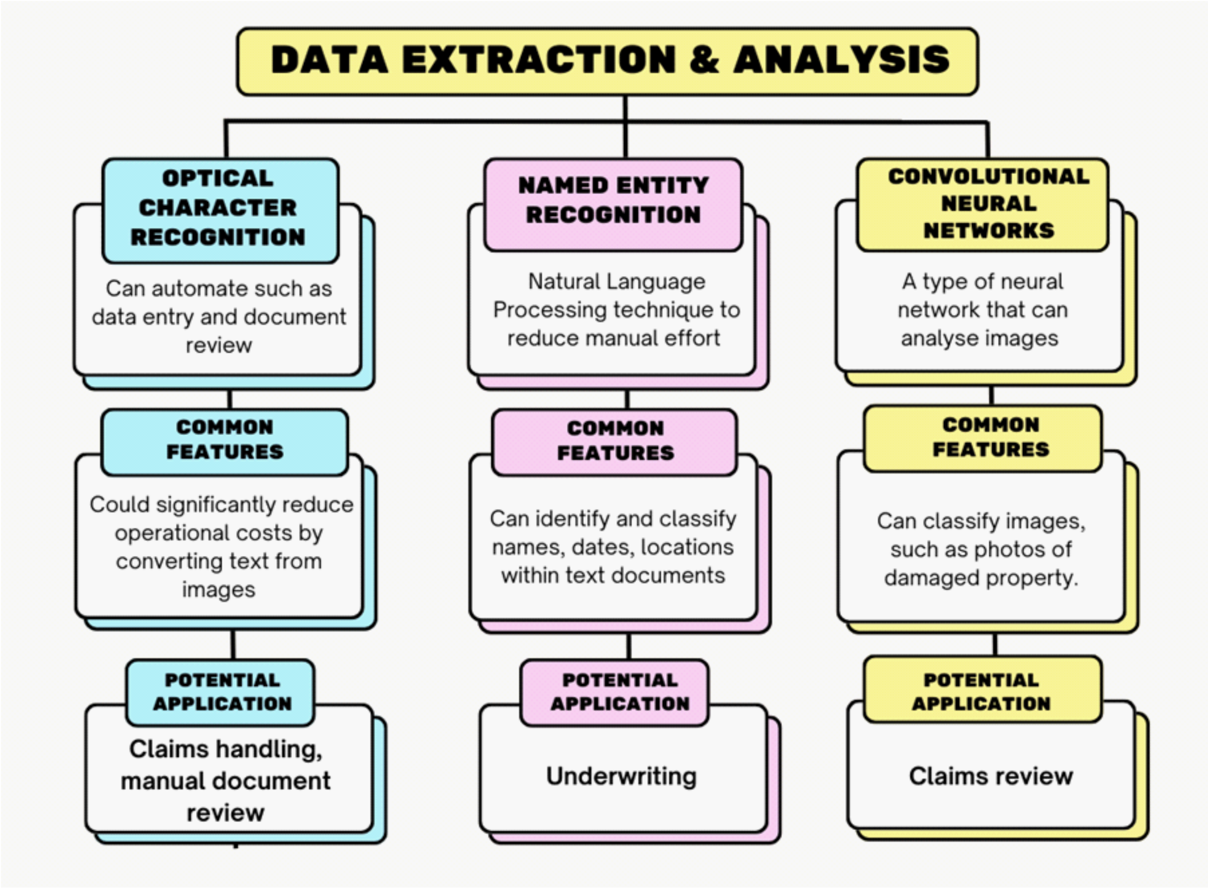
A chain-ladder model for claims reserving was extended to include individual claims feature information. This extension was done by using a neural network model for the factors. This extended model had the advantage that we could analyse reserves on individual claims, which allows us, for instance, to capture changing portfolio mixes.

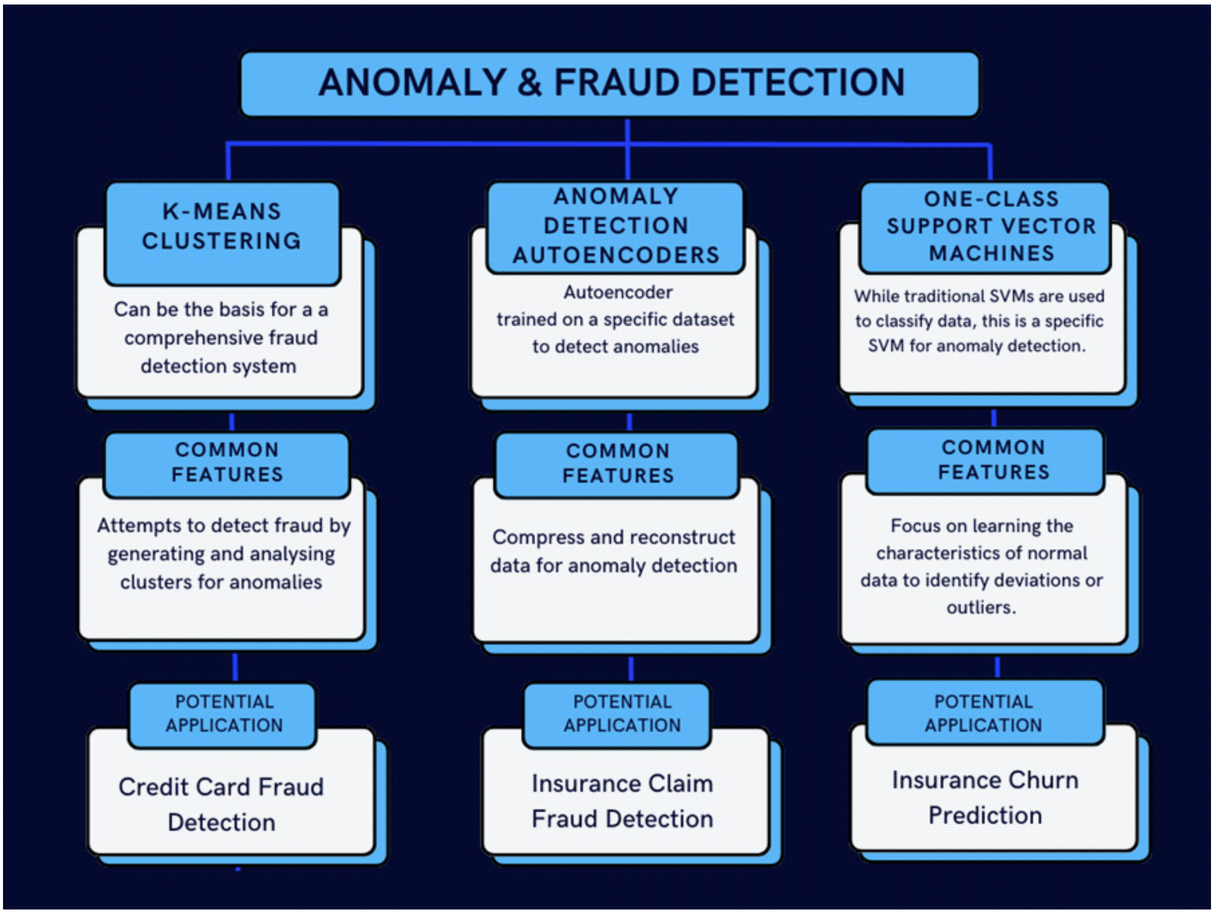
The work was only considered a starting point of a much broader modelling task, as it only considered static feature information, i.e., the feature labels do not change over time. The modelling task is expected to be more complex for dynamic feature components, because dynamic features will require modelling of multidimensional stochastic processes.

[44] Source: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8007210
[45] Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2966126

APPENDIX B: SAMPLE SOLUTIONS




DATA EXTRACTION & ANALYSIS
OPTICAL CHARACTER RECOGNITION
Can automate such as data entry and document review
NAMED ENTITY RECOGNITION
Natural Language Processing technique to reduce manual effort
CONVOLUTIONAL NEURAL NETWORKS
A type of neural network that can analyse images
COMMON FEATURES
Could significantly reduce operational costs by converting text from images
COMMON FEATURES
Can identify and classify names, dates, locations within text documents
COMMON FEATURES
Can classify images, such as photos of damaged property.
POTENTIAL APPLICATION
Claims handling, manual document review
POTENTIAL APPLICATION
Underwriting
POTENTIAL APPLICATION
Claims review


ANOMALY & FRAUD DETECTION
K-MEANS CLUSTERING
Can be the basis for a a comprehensive fraud detection system
ANOMALY DETECTION AUTOENCODERS
Autoencoder trained on a specific dataset to detect anomalies
ONE-CLASS SUPPORT VECTOR MACHINES
While traditional SVMs are used to classify data, this is a specific SVM for anomaly detection.
COMMON FEATURES
Attempts to detect fraud by generating and analysing clusters for anomalies
COMMON FEATURES
Compress and reconstruct data for anomaly detection
COMMON FEATURES
Focus on learning the characteristics of normal data to identify deviations or outliers.
POTENTIAL APPLICATION
Credit Card Fraud Detection
POTENTIAL APPLICATION
Insurance Claim Fraud Detection
POTENTIAL APPLICATION
Insurance Churn Prediction

REFERENCES

1. Richman, Ronald (2018). AI in Actuarial Science. Paper presented at the 2018 ASSA Convention, Cape Town. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3218082 [accessed December 2023]
2. Wuthrich, Mario V. & Buser, Christoph (2023). Data Analytics for Non-Life Insurance Pricing. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2870308 [accessed December 2023]
3. Pouget, Fanny et al (2021). The use of artificial intelligence and data analytics in life insurance. https://www.milliman.com/en/insight/the-use-of-artificial-intelligence-and-data-analytics-in-life-insurance [accessed December 2023]
4. Richman, Ronald (2021). Mind the gap – safely incorporating deep learning models into the actuarial toolkit. Published in British Actuarial Journal (2022), **Vol. 27**, e21, pp. 1–57. https://deliverypdf.ssrn.com/delivery.php [accessed December 2023]
5. Panlilio, Alex et al (2018). Practical Application of Machine Learning with Actuarial Science. https://www.actuaries.org.uk/system/files/field/document/Practical%20Application%20of%20Machine%20Learning%20within%20Actuarial%20Work%20Final%20%282%29_feb_2018.pdf [accessed December 2023]
6. Wuthrich, Mario V. & Merz, Michael (2023). Book titled Statistical Foundations of Actuarial Learning and its Applications. https://link.springer.com/book/10.1007/978-3-031-12409-9 [accessed December 2023]
7. What Data Science means for the future of the Actuarial Profession – Abstract of the London Discussion (2018). Published online by Cambridge University Press. Published in British Actuarial Journal, **Vol. 23**, e16, pp. 1–21. © Institute and Faculty of Actuaries 2018. https://www.cambridge.org/core/journals/british-actuarial-journal/article/what-data-science-means-for-the-future-of-the-actuarial-profession/B42EF98A53ED8D6E42606FAE664021D2 [accessed December 2023]
8. Lorenzoni, Andrea & Reilly, Michael (2022). Why AI in insurance claims and underwriting? https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-Why-AI-In-Insurance-Claims-And-Underwriting.pdf [accessed December 2023]
9. Chui, Michael et al (2022). The State of AI in 2022 – and a half decade in review. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review [accessed December 2023]
10. 2021 Insurance Factbook (2021). Published by Insurance Information Institute. https://www.iii.org/sites/default/files/docs/pdf/insurance_factbook_2021.pdf [accessed December 2023]
11. IBM Corporation. Cost of a Data Breach Report 2022 (2022). https://www.ibm.com/downloads/cas/3R8N1DZJ [accessed December 2023]

CASE STUDY 1

Dutta S, Bandyopadhyay SK (2020). Detection of Fraud Transactions Using Recurrent Neural Network during COVID-19. https://www.researchgate.net/publication/347478467_Detection_of_Fraud_Transactions_Using_Recurrent_Neural_Network_during_COVID-19 [accessed January 2024]

CASE STUDY 2

Martini, Antonio (2019). Deep Recurrent Neural Networks for Fraud Detection on Debit Card Transactions . https://www.crc.business-school.ed.ac.uk/sites/crc/files/2020-10/E29-Deep-Recurrent-Neural-Networks-Martini.pdf [accessed January 2024]

CASE STUDY 3

Fashoto, Stephen Gbenga et al (2016). Hybrid Methods for Credit Card Fraud Detection Using K-means Clustering with Hidden Markov Model and Multilayer Perceptron Algorithm.

https://www.researchgate.net/publication/287358941_Hybrid_Methods_for_Credit_Card_Fraud_Detection_Using_K-means_Clustering_with_Hidden_Markov_Model_and_Multilayer_Perceptron_Algorithm [accessed January 2024]

CASE STUDY 4

Ripan, Rony Chowdhury et al (2020). An Isolation Forest Learning Based Outlier Detection Approach for Effectively Classifying Cyber Anomalies.

https://www.researchgate.net/publication/348380019_An_Isolation_Forest_Learning_Based_Outlier_Detection_Approach_for_Effectively_Classifying_Cyber_Anomalies [accessed January 2024]

CASE STUDY 5

Oracle.com. What Is Natural Language Processing (NLP)? https://www.oracle.com/artificial-intelligence/what-is-natural-language-processing/ [accessed January 2024]

Lemonade.com. Lemonade's Year in Review, 2017. https://www.lemonade.com/blog/lemonade-review-2017/ [accessed January 2024]

Lemonade.com. Two Years of Transparency: A Super Transparency Chronicle. https://www.lemonade.com/blog/two-years-transparency/ [accessed January 2024]

Lemonade.com. The Sixth Sense: Lemonade's 2019 Product in Review: Blond wigs, weird bots, and ad fails. https://www.lemonade.com/blog/the-sixth-sense/ [accessed January 2024]

CASE STUDY 6

Cote, Marie-Pier et al (2020). Synthesizing Property & Casualty Ratemaking Datasets using Generative Adversarial Networks. https://hartman.byu.edu/docs/files/CoteHartmanMercierMeyersCummingsHarmon_SynthesizingDataGANs.pdf [accessed January 2024]

CASE STUDY 7

Martin, Laurene (2020). Mortality Risk Modelling with Machine Learning. https://www.institutdesactuaires.com/docs/mem/a1c834b2b50a8592bfc120173a0c1db2.pdf

[accessed January 2024]

CASE STUDY 8

Wuthrich, Mario V. & Merz, Michael (2023). Book titled Statistical Foundations of Actuarial Learning and its Applications. https://link.springer.com/book/10.1007/978-3-031-12409-9 [accessed January 2024]

CASE STUDY 9

Eriksson, Alexander. A Comparison of Gradient Boosting Machines and Generalized Linear Models for Non-Life Insurance Pricing. Mathematical Statistics Stockholm University Master Thesis. http://www.math.su.se [accessed January 2024]

CASE STUDY 10

Lai, Dihui (2018). Wired to Underwrite: Artificial Intelligence and Underwriting. https://www.rgare.com/knowledge-center/article/wired-to-underwrite-artificial-intelligence-and-underwriting [accessed January 2024]

CASE STUDY 11

Van Ruitenbeek, R.E. and Bhulai, S. (2022). Convolutional Neural Networks for vehicle damage detection. https://www.sciencedirect.com/science/article/pii/S2666827022000433 [accessed January 2024]

CASE STUDY 12

Lin, Weiwei et al (2017). An Ensemble Random Forest Algorithm for Insurance Big Data Analysis. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8007210 [accessed January 2024]

CASE STUDY 13

Wuthrich, Mario V (2017). Neural Networks Applied to Chain-Ladder Reserving. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2966126 [accessed January 2024]

DATA, DISRUPTIONS AND THE ACTUARY

12, 13, 14 FEBRUARY, 2024 AT THE WESTIN POWAI LAKE, MUMBAI

PAPER II

The importance and impact of digital trust in the era of data and AI

The importance and impact of digital trust in the era of data and AI

Mitali Chatterjee
Nikhilmon O U
Simon Woodward

ABSTRACT:

The digital world is increasingly blending with the physical world. Digital technologies, data and analytics are playing an important role in helping insurers close protection gaps around the world, assess climate change related risks and supply chain vulnerabilities and analyze mortality and morbidity trends. For example, economies with greater digital penetration are typically more resilient to other exposures such as natural catastrophes. However, digital transformation is incomplete without establishing digital trust. In this paper, we show how digital trust can generate positive externalities for trade, innovation and entrepreneurship, macroeconomic performance, social wellbeing, and cyber resilience. Lack of digital trust and digital misinformation, on the other hand, can be costly for the economy, resulting in loss to national income. This motivates us to dig deeper to analyze factors that can potentially influence digital trust. Results of our quantitative analysis reveals that three of the nine factors identified are relatively more important dimensions for decoding digital trust. These include cultural and generational attitudes, ease of use and explicability of artificial intelligence/ automated decision making. We also perform a cross sectional regression to show that high digital trust translates into better insurance density while an ensemble machine learning approach using random forest model reveals that high digital trust contributes to better insurance penetration.

JEL classification: G22, 033, C45, E70

KEYWORDS:

Digital, trust, insurance, artificial intelligence

[1] This paper has been submitted for the 23rd Global Conference of Actuaries (GCA) 2024
[2] Core discipline: Data Sciences and Analytics
[3] The information is generic in nature and should not be considered in the form of professional advice. Swiss Re does not guarantee the accurateness and completeness of the information. Swiss Re shall not be liable for any damages arising out of or in connection with this document. Unless otherwise stated or sourced, views of the authors are largely personal and not reflective of their affiliation to any organization.
[4] Mitali Chatterjee is an Insurance Research Manager at Swiss Re Institute. Her key focus areas include digital ecosystem, mental health, public sector solutions, sustainability and decarbonization. Prior to joining Swiss Re, Mitali has held several thought leadership roles across organizations and authored several papers focusing on cross-cutting emerging technologies and their impact on the public sector. By education, she is an economist with a Bachelor's degree from St Xavier's College, Kolkata and a Master's degree from Indira Gandhi Institute of Development Research (IGIDR).
[5] Nikhilmon O U is an insurance senior research analyst, based at Swiss Re Institute. As a part of thought leadership and market intelligence team, he primarily focuses on research and analytics related to decarbonization of critical economic systems. Prior to joining Swiss Re, Nikhil has worked with several bilateral and multilateral agencies in project design and strategy development, research, and monitoring and evaluation across different development sectors. Key interest areas include climate change, low carbon economy, regenerative agriculture, livelihoods, natural resources management, gender, and humanitarian issues.
[6] Simon Woodward is a senior content developer & editor at Swiss Re Institute. Simon graduated from the Universities of Birmingham and Oxford before working in political and sovereign risk in London. Moving to Zurich, he has worked in various functions in the reinsurance industry, including as an editor, a researcher, in communications, in event content development and with the Swiss Re Strategic Council.
[7] The authors would like to thank Jonathan Anchen, Head Market Intelligence, Swiss Re Institute for his inputs on the draft of this paper.

CHAPTER 1: BUILDING THE DIGITAL TRUST PYRAMID[8]

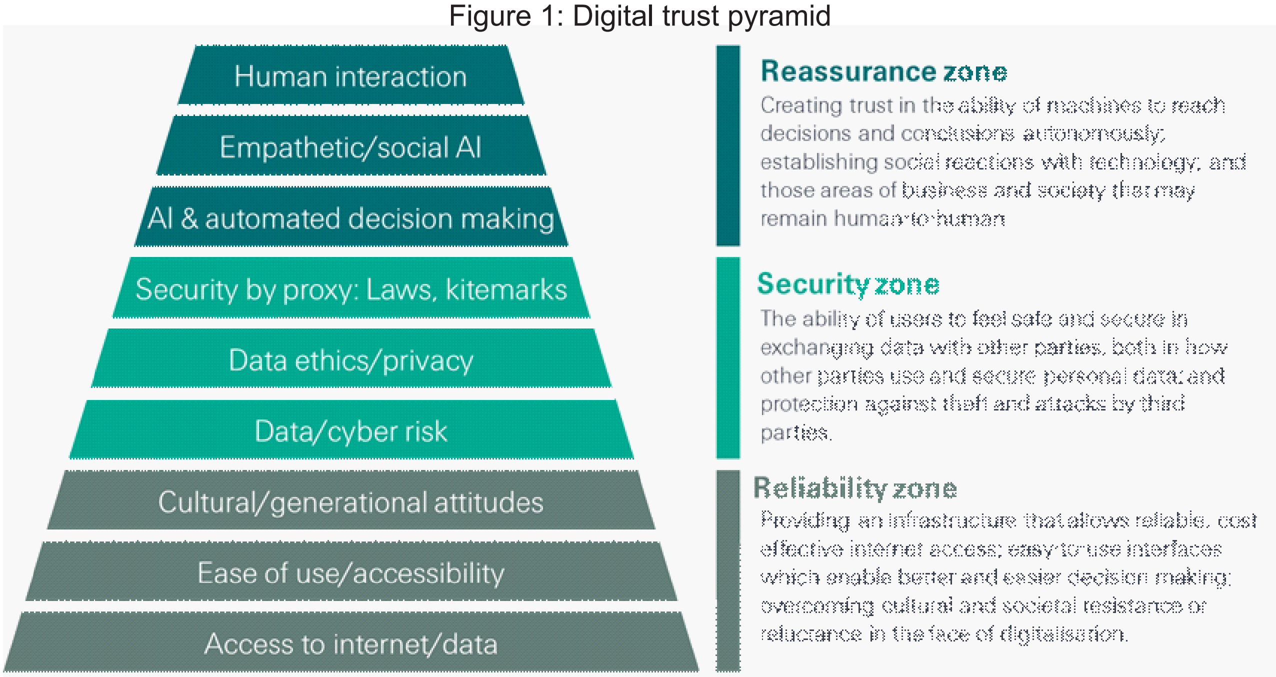
The concept of digital trust is amorphous and subjective. There is no single data point that can define digital trust. It is not a commodity to be bought, not an asset that can be traded, not a risk that can be underwritten. If trust has a value, then it is in the perceived strength of the relationship between the individual, business or institution and the party with which it seeks to interact. The notion of trust is, to a large extent, subjective and mutable, making it difficult to measure.

There have been a number of attempts to capture definitions of digital trust. The World Economic Forum defines it as the expectation of individuals that their digital technologies "will protect all stakeholders' interests and uphold societal expectations and values.[9]" The Digital Trust Label defines digital trust as a product of four components, security, data protection, reliability and fair user interaction.[10] Consultancy McKinsey regards digital trust as "consumer faith in cybersecurity, data privacy, and responsible AI[11]." Rather than take a top-down approach, focused either through the prisms of business, technology or society, we have chosen a bottom-up individualised approach with a more psychological bias, constructed in-house by Swiss Re Institute, loosely inspired by Abraham Maslow's pyramid of needs from the 1943 publication "A Theory of Human Motivation." We believe this might help the insurance industry with understanding their clients.

For an insurer, the march towards digital can be transformative across the business. It provides a point of interaction with clients; it can automate practically all processes across the insurance customer journey; and it can enhance understanding of our portfolios and can optimise how we cover risk. Underpinning these processes of digitalisation is trust: trust from our employees, trust from our regulators, our stakeholders, our shareholders and most importantly, our customers.

We thus need a means of conceptualising and understanding what trust means in a digital context. One method of doing this is the digital trust pyramid.

Figure 1: Digital trust pyramid



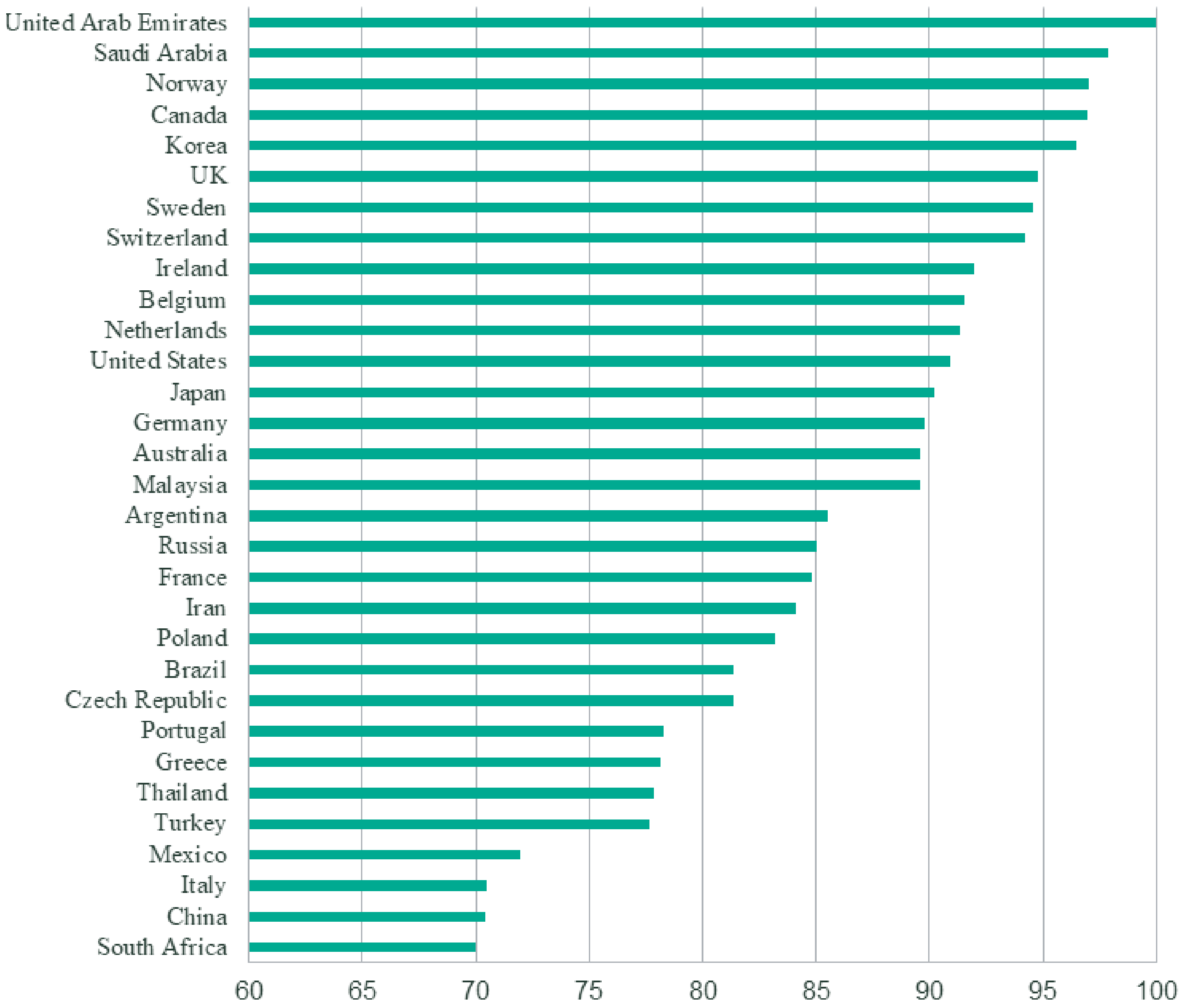
Source: Decoding digital trust – An insurance perspective, Swiss Re Institute (SRI)

[8] For a detailed understanding, please refer to Decoding digital trust – An insurance perspective, Swiss Re Institute.
[9] initiatives.weforum.org/digital-trust/about
[10] www.swiss-digital-initiative.org/digital-trust-label/
[11] www.mckinsey.com/capabilities/quantumblack/our-insights/why-digital-trust-truly-matters

Reliability zone

This pyramid is constructed of nine building blocks within three zones. The first of these zones **– reliability –** is largely mechanistic. It seeks to understand the physical and mental constraints of digital trust. Digital relationships cannot be established without consistent and reliable data and internet access. We find that internet penetration to be closely correlated with both wealth and economic development in most cases. Developing economies typically have lower internet penetration rates - only one sub-Saharan African state, for example, reaches 70% internet penetration (Figure 2).

Figure 2: Internet penetration, selected states, % 2020



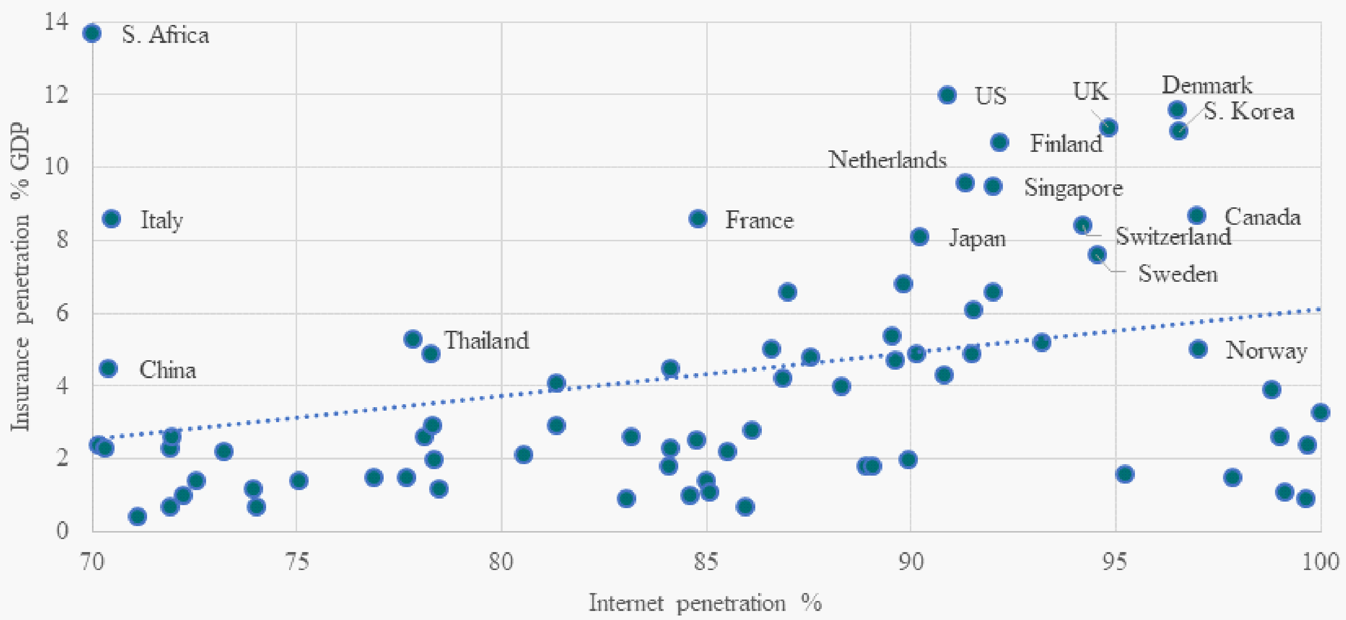
Source: ITU[12]

[12] Decoding digital trust – An insurance perspective, Swiss Re Institute

Access to the internet is not a given. It is dependent on fixed and mobile infrastructure as well as the costs of accessing data, both factors which will influence the number of users and amount of internet usage. The Inclusive Internet Index of the Economist Intelligence Unit provides an extensive scoring methodology for selected countries, based on both the supply and demand sides.[13] In terms of affordability, Canada sits at the top of the 2020 rankings, followed by the UK, France and Italy. Smaller developed economies including Singapore, Hong Kong, South Korea, Switzerland and Denmark dominate the top of the availability rankings.

The reason why it is important to think about internet access from an insurance industry standpoint is because insurance penetration and internet penetration are positively correlated, as we can see from Figure 3. The associations of this correlation are multiple. Income will be a strong underlying factor bringing the two together and more visible among lower income countries (with some outliers). However, insurance penetration has a number of drivers, including on the supply side regulatory support and development of the financial sector; and on the demand side, the culture of risk coverage, the role of the state in risk mitigation, income distribution and the specific risk exposure of a particular country, for example natural catastrophes.

Figure 3: Insurance & Internet penetration, selected countries, 2020



Source: ITU and SRI [14]

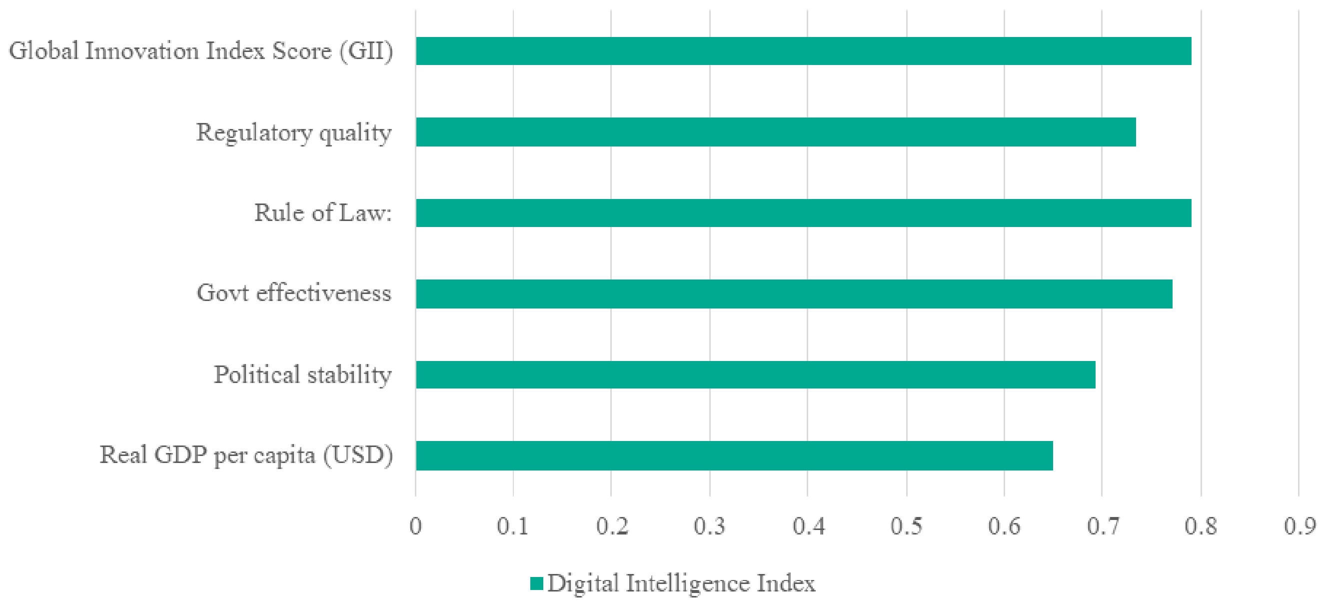
Access to the internet will only be advantageous if it allows the user access to services they could not otherwise receive; or easier, more convenient access to goods and services they could get elsewhere. In essence, the question is one of ease and, specifically, how easy it is to use your digital portal, at least, on the supply side. The demand side also plays a role. We check if there are cultural, age or other societal factors at play that influence our willingness to trust the digital interface.

Societies may have certain inherent characteristics that either fuel or inhibit digital trust. These could be linked to political structures, socio-economic factors, cultural identities or other values. Our analysis (Figure 4) reveals that countries with higher levels of digital trust usually exhibit one or more of the following cultural characteristics: they tend to have more innovative populations, enjoy higher incomes, and exhibit better governance.

[13] The Inclusive Internet Index, The Economist, 2021.

[14] Decoding digital trust – An insurance perspective, Swiss Re Institute

Figure 4: Correlation between digital trust and country characteristics



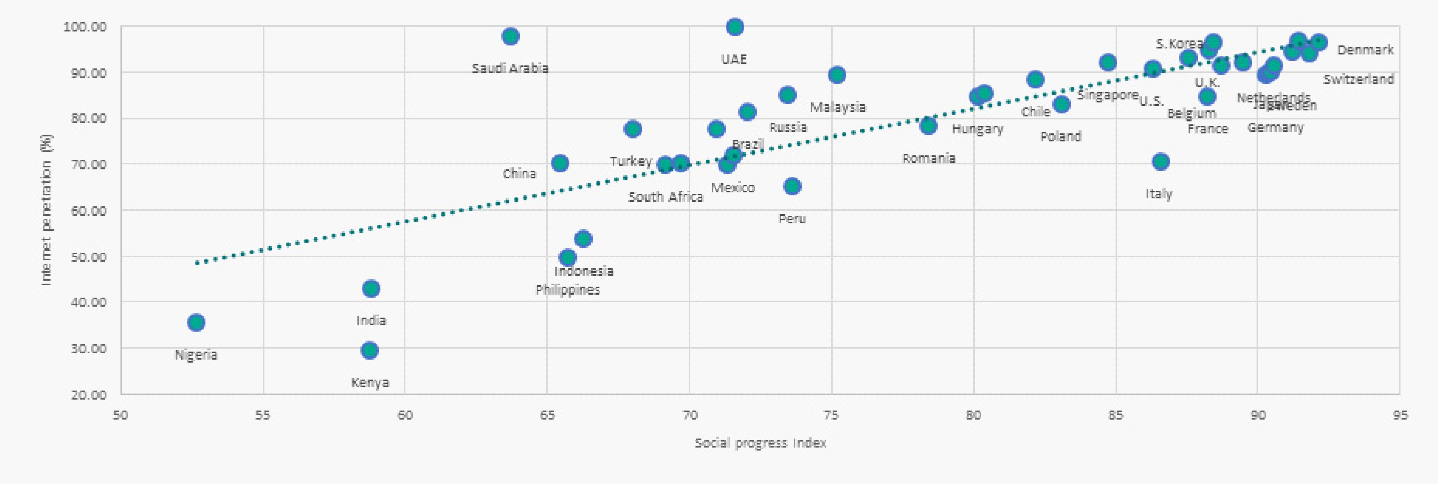
Source: SRI, Tufts University, World Bank and GII[15]

Countries that have made greater strides socially are usually willing to trust the digital space more. Using the Digital Intelligence Index (DII) published by the Tufts University and the Social Progress Index 2021, we find that there is a moderately high degree of correlation between the two indices (0.61).[16], [17]Social progress is also positively associated with the degree of internet penetration (correlation coefficient of 0.88) - countries with a high degree of social progress have a greater internet penetration. Figure 5 highlights this association. Some outliers (such as Italy) highlight our earlier point that digital access and trust do not track against any single variable. Progress on social issues does not automatically accompany economic development. While the Nordic countries are some of the best performers in the SPI, countries like Fiji, Sri Lanka and the United Arab Emirates have recorded the most significant progress in the last ten years.[18]

Figure 5: Greater internet openness and digital trust have positive externalities for trade, innovation and entrepreneurship, macroeconomic performance, and social wellbeing.[19]

[15] Decoding digital trust – An insurance perspective, Swiss Re Institute
[16] Digital in the time of COVID, Tufts University, 2019.
[17] The Social Progress Index (SPI) takes into consideration 53 social and environmental indicators across three themes - basic human needs, foundations of wellbeing, and opportunity to measure progress of nations on the quality of lives of their citizens. See Global Index: Overview, socialprogress.org, 2021.
[18] 2021 Social Progress Index, Social Progress Imperative, 2021.
[19] 2021 Social Progress Index, Social Progress Imperative, 2021.

Figure 5: Positive correlation between social progress and internet penetration



Source: SRI, ITU, and Social Progress Imperative[20]

Cultural and generational attitudes are also an important determinant of how citizens choose to interact with local and national governments. With changes in demographic patterns, urbanisation and technological advances, there is a marked shift from offline to online engagement between the state and its citizens in many countries. The extent to which such a transition is successful largely depends on the historical attitudes of a country's population towards change and innovation and the perceived trust levels of that population in both the government and digital platforms.

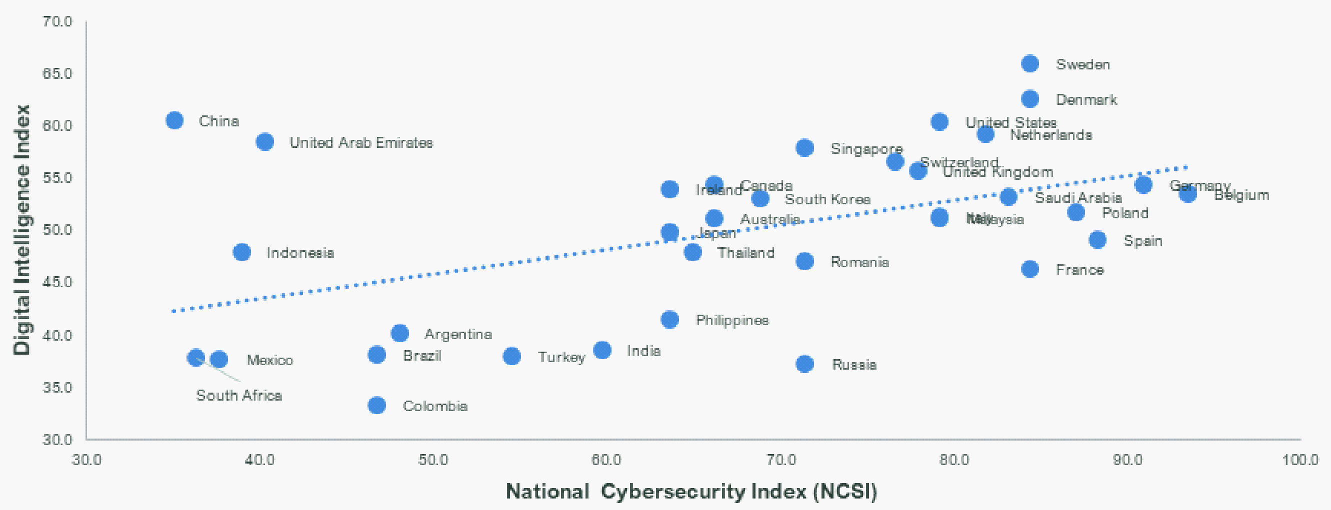
SECURITY ZONE

Our second zone is that of **security**. We assume that access, ability and willingness are already there for a digital relationship to blossom. But how do companies maintain that trust as the relationship develops? Our starting point is security and the forceable intrusion of third parties – cyber risk in its many forms, from identity theft to blackmail, from data hijacking to digital larceny.

Cyber resilience is an important step towards establishing digital trust. SRI analysis finds that there is a strong positive correlation between cyber preparedness and digital trust, suggesting that countries that invest in cybersecurity policies, legislation and outcomes do experience greater levels of digital trust within their economies. Figure 6 highlights this cross-sectional relationship.

[20]Decoding digital trust – An insurance perspective, Swiss Re Institute

Figure 6: Positive association between digital trust and cyber resilience



Source: SRI, Tufts University, and e-Governance Academy Foundation[21]

With security from attack assured, clients and customers will focus on fresh concerns: they do not want their digital trust abused; their data used without consent; or the storage of data outside of contractual obligation. Cisco's Consumer Privacy Survey in June 2020 revealed that 31% of respondents were concerned that their data would be used for unrelated purposes, while 25% were worried that their data may be shared with third parties.[22]

We also deliberate on the role of legislation in forming trust. Widely regarded as best-in-class is the EU's General Data Protection Regulation (GDPR). Implemented in 2018, and running to 99 articles, the GDPR's focal point is personal data. Companies should hold no more data than necessary and cyber security should be appropriate to the size of the organization.[23] The US, by contrast, lacks a comprehensive federal data protection law. One might expect a good correlation between digital trust and the strength of digital security legislation- however, existing literature is divided in its opinion. While no regulation can mean little trust, excessive legislation may result in erosion of trust.

REASSURANCE ZONE

Lastly, we reach the **zone of reassurance**. The future is inherently uncertain. We have to make decisions today that will affect us tomorrow. Will we let technology – in the form of artificial intelligence – make these decisions for us? Looking through the literature, two themes quickly coalesce around AI. The first is game-changing. Across a wide vista of activities, the suggestions are that AI will be transformative. It has the potential to create efficiencies, enable discoveries and facilitate automation. This is very much the case within the insurance industry. As IBM predicted in 2021: "In the coming years, automation and AI solutions will roll out across every domain in the insurance industry".[24] Current predictions expect AI will be able to provide insurance customers with a more personalised journey across the insurance value

[21] Decoding digital trust – An insurance perspective, Swiss Re Institute
[22] Cisco June 2020 survey of more than 2600 respondents across 12 countries- See Protecting Data Privacy to Maintain Digital Trust, Cisco, 2020.
[23] M. Burgess, "What is GDPR? The summary guide to GDPR compliance in the UK", Wired, 24 March 2020.
[24] From underwriting to claims management, artificial intelligence will transform the insurance industry, IBM, 13 September 2021.

chain, while empowering insurers to perceive new insights into both their customers and the risks their customers face. AI should allow insurers to overcome the inefficiencies of legacy systems, consolidate consumer data and breach previously held silo walls.

An IPSOS survey for the World Economic Forum in 2022 explored attitudes to AI in 28 countries.[25] It reported a positive correlation between a perceived understanding of AI among respondents and trust in companies using the technology. However, understanding may not have been the only driver of trust. Both levels of trust and levels of understanding were higher among emerging markets than developed markets. Figures for trust and understanding of AI were significantly lower, for example, in Japan – hardly a stranger to digital technology – than India or China, which suggests cultural factors at play. Also distrustful of companies using AI were respondents from Germany, Canada, the US and the UK.

The jump between AI and emotionally literate or sensitive AI is a significant one. Smart machines can learn from data sets. The data set of human interactions is a hugely granular and complex undertaking. Moreover, human interactions are executed through many levels of sensual perception, including language, facial response, body language and even smell. To become more challenging still, human relationships are also undertaken through the lens of particular languages and cultures. The field is an established one, dating back at least to the 1990s and the Affective Computing Group at MIT Media.[26]

The point we are trying to drive is that consumers will show greater trust in empathetic AI, tempered by cultural and possibly generational perspectives. However, emotional training of AI currently consumes time and investment and there are concerns that providing a service that becomes overly realistic but not-quite-human could conversely erode trust.

Finally, we want to address the notion that that digitalization, and AI, could bring sweeping changes to the world of white-collar work. Processes currently undertaken by professionals, in fields such as health, law and financial services, would be increasingly automated, rendering many jobs, as we currently understand them, obsolete. At the same time, studies suggest complementary effects of technology on labor and the ability of humans to provide problem-solving skills, adaptability, and creativity skills.

Similarly, building an effective digital trust strategy also necessitates understanding where human trust may be necessary and complementary. For example, in health human reassurance is most important- one would not want to hear the news on malignant tumor from a robot. Similarly, in aviation, people may be hesitant in boarding a flight that doesn't have a human pilot, even though in fact, 90% of a typical flight is already performed by autopilot. Absence of digital trust, driven by factors such either insufficient regulation and oversight, lack of ethics, or misuse of AI can result in spread of misinformation and further mistrust that can prove to be costly.

To conclude this section, digital trust is a subset of the wider notion of trust. The factors that affect trust across nations, cultures, language groups, class, genders - indeed any subdivisions of people - are numerous and not easily reconciled. While this chapter tries to qualitatively build on the factors driving digital trust, the next chapter tries to quantify the importance of each of the 9 blocks making the digital trust pyramid. Finally, in Chapter 3 we build a digital trust score for 90 countries and check whether digital trust is an important factor in determining insurance penetration and density.

[25] J. Myers, "5 charts that show what people around the world think about AI", World Economic Forum, 5 Jan 2022.

[26] M. Sommers, "Emotion AI, explained", MIT Management Sloan School, 8 March 2019.

Box 1: Absence of digital trust results in misinformation, which is costly.

Where the internet is available, it is available – barring censorship - to all. It creates a 'too-much-too-little' scenario - too much information is available, but only a small proportion may be relevant and accurate. The Reuter's Institute Digital News Report 2021 shows that while 82% of survey respondents read news online (including social media), only 34% trust search engines for news whilst 24% trust the news they read on social media.[27]

We can estimate the quantitative cost of misinformation. To do so, we took two case studies of climate change and non-communicable diseases (NCDs). We used a funnel approach of a total population of 18 countries using social media as their primary news source. We then estimated the share of social media users with an interest in climate change and the probability of those viewing inaccurate or misinformation on climate change. Studies suggest that the economic cost of climate change per person (accounting for both cost to humanity as well as impact of climate change on economic growth) could be as high as USD15,000.[28] SRI analysis revealed that digital misinformation-driven climate change inaction could cost the world **USD37.5 billion**, or roughly **0.05% of global GDP**.[29]

NCDs are driven by poor diet, nutrition, and lack of exercise. They result in 41 million premature deaths annually.[30] We wanted to estimate the mortality and morbidity cost associated with digital misinformation on diet, nutrition, and exercise. Research shows that almost 60% of younger consumers refer to digital media for news related to health and nutrition. Further analysis reveals that up to 40% of the most frequently shared links often contain fake news. Using a combination of this data along with figures on NCD years of life lost (YLLs)[31] from the Global Burden of Diseases database, we arrived at an estimate for NCD YLLs due to misinformation: 26.26 million years of lives lost in the selected sample of countries. Using per capita income as a proxy for per capita YLL cost for the same set of 18 countries, we arrived at the huge figure of **USD 962.81 billion** in costs to the world (or **1.1% of global GDP**) resulting directly from digital misinformation relating to the premature mortality cost of NCDs.[32]

These findings, combined with a plethora of research associating misinformation with the COVID-19 pandemic, suggests that digital trust, reliability, and authenticity come not only with significant economic but also human costs.

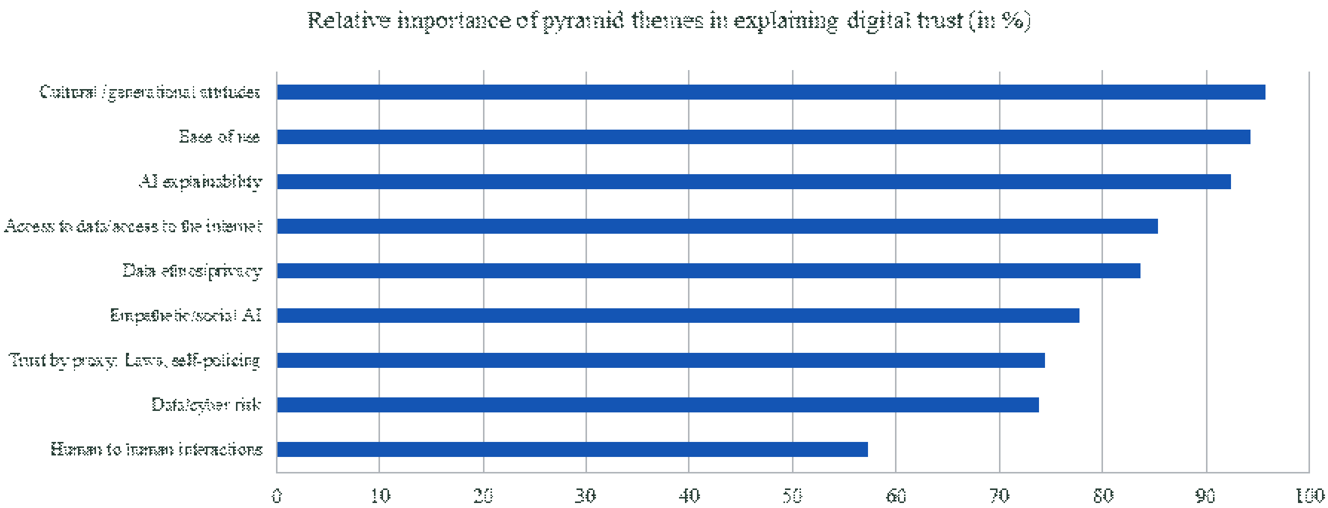
[27] Digital News Report 2021, Reuters Institute, 2021.
[28] Economic cost of climate change could be six times higher than previously thought, UCL News, 6 September 2021.
[29] Refer to Appendix.
[30] Non-communicable Diseases, World Health Organization, 13 April 2021.
[31] The YLLs for a cause are calculated as the number of cause-specific deaths multiplied by a loss function specifying the years lost for deaths as a function of the age at which death occurs
[32] Refer Appendix 1

CHAPTER 2: DIGITAL TRUST PYRAMID DIMENSIONS - DIGGING DEEPER[33]

Our process to quantify the impact of each of dimension on levels of digital trust begins by using four to six alternative parameters as proxies for each of the nine dimensions (total of 46 parameters), for a total of 120 countries, and assess their robustness. For example, access to data can be measured by parameters such as individuals using the Internet (% of population), fixed broadband subscriptions (per 100 people), and share of population with 4G mobile network coverage. We can subsequently check on feasibility and data consistency to narrow our analysis down to the most suitable signals.[34] For more on the parameters considered, see Appendix 2.

For the parameters included within each dimension, we create a weighted average index for each of the nine dimensions for all the countries analysed. We then use these standardised country specific dimensional scores to group countries into clusters.[35] We created four clusters of countries with similar digital trust qualities (see Figure 8). Figure 7 below is an analysis of variance demonstrating the relative ranking of dimensions that have the maximum ability to separate country to clusters. The purpose of this exercise is to identify the factors that have the highest explanatory power in determining digital trust, globally. Though all the nine dimensions are important, some may fuel accelerated growth in digital trust. Our analysis reveals that the three most important dimensions are cultural and generational attitudes, ease of use and explicability of AI/ automated decision making. The rest of the chapter investigates how/ why these three themes are critical in decoding digital trust.

Figure 7: Relative importance of digital trust themes[36]



Source: SRI[37]

[33] For more details, please refer to Digital trust II: A consumer perspective, Swiss Re Institute

[34] We use the Cronbach's alpha method to check for internal consistency - dropping variables that don't have high covariance relative to variance. Please refer to Appendix 2 for detailed explanation

[35] K-means clustering is an unsupervised learning algorithm that groups an unlabelled dataset into a number of pre-defined clusters: in this analysis we are trying to cluster countries that behave similarly as opposed to others

[36] Refer Appendix 2 for detailed methodology and statistical tables

[37] Digital trust II – A consumer perspective, Swiss Re Institute

Cultural/ generational attitudes

Cultural and societal attitudes rank highly in our analysis with an ability to contribute up to 95% of variance in digital trust between country clusters. Scandinavian countries outperform in terms of digital trust scores. The reason may be rooted in their history and culture.[38] These countries have historically displayed higher levels of interpersonal trust and trust in their institutions (including when they collect data). This may also translate into higher levels of digital trust. They also have deep belief in transparency, social and environmental responsibility and are therefore more willing to lend digital trust to businesses and institutions that are perhaps driven by a sense of purpose, and not profit alone. Countries with high levels of societal trust tend to do better with digital trust.

Cultural barriers on the other hand may impede digital trust and societal cooperation. Such barriers include an inherent risk aversion, fear of change, hierarchical decision-making, and political and institutional inertia.[39] Research suggests that cultural risk aversion and fear of change in countries is closely associated with historical shock events such terrorism and natural disasters.[40] People in an earthquake prone country like Indonesia are traditionally risk averse and may be less amenable to digital trust.[41]

Generational attitudes make a difference as well. For example, last year a consumer survey by Swiss comparison site Comparis assessed the importance of incentivisation in the disclosure of health-related information.[42] At the time, 9% of respondents said they were happy to disclose health-related information to their insurer online. A quarter of total said they would do so in return for a reward. Around 5% would be happy to share personal data for a monthly reward of CHF 5; a further 12% would want at least CHF10; and 19% would disclose information for CHF 20 or more. A further 34% said they would share their data for a monthly reward of CHF 50 or more. Most resistance came from the 50 to 65-year-old cohort, with 46% saying they would not provide information at any price. That figure fell to 25% among those aged 30 to 49, and to 17% for the under-30s.

While it is hard to reach a global generalisation using such anecdotal examples, these do help support our findings in favour of cultural and generational factors playing an important role in explaining digital trust.

EASE OF USE

Ease of use is the second most important factor according to our variance analysis, explaining more than 94% of the variation in digital trust. The importance of perceived ease of use in strengthening digital trust is well documented. Digital platforms should allow for ease of functionality, navigation and search, along with features such as portability and interoperability. One study finds of use as measured by ease of recognition, ease of navigation, ease of obtaining information and ease of purchase as positively influencing digital trust in e-banking platforms.[43] This includes having an operating system that is intuitive to use and designing a digital platform where the fine print of the terms and conditions can be easily read and understood. The study was consumer survey-based with questions on the use of banking apps.

[38] de Godoy, Jaqueline, Kathrin Otrel-Cass, and Kristian Høyer Toft, "Transformations of trust in society: A systematic review of how access to big data in energy systems challenges Scandinavian culture", Energy and AI, 2021.
[39] Wilson, Christopher, and Ines Mergel, "Overcoming barriers to digital government: mapping the strategies of digital champions" Government Information Quarterly, vol 39.2, 2022.
[40] Cicerale, Alessandro, Enrico Blanzieri, and Katiuscia Sacco, "How does decision-making change during challenging times?" PLoS One, 2022.
[41] Thamarapani, Dhanushka, and Marc Rockmore, "The stability and evolution of risk attitudes and time preferences after a disaster", International Journal of Disaster Risk Reduction, vol 70, 2022.
[42] Fitness trackers and apps – Health insurance, Comparis, 2022.

[43] Martínez-Navalón, Juan-Gabriel, María Fernández-Fernández, and Fernanda Pedrosa Alberto, "Does privacy and ease of use influence user trust in digital banking applications in Spain and Portugal?", International Entrepreneurship and Management Journal, 2023.

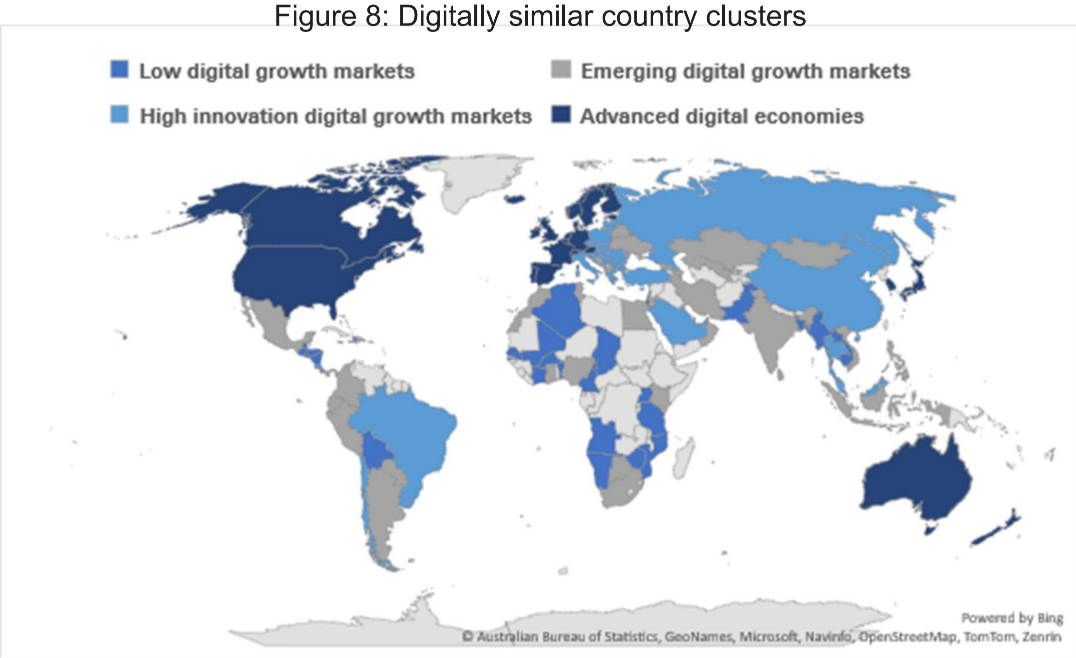
AI/ AUTOMATED DECISION MAKING

As above, the increasing fusion of digital and real worlds fuelled by sensors and AI is changing the parameters influencing digital trust. Figure 3 shows that AI/automated decision making is also an important factor contributing to the variation in digital trust. However, AI models are often a black box for consumers. They are unable to see the benefits of AI as an intelligent decision-making tool that can perform complex computational task unless they understand the mechanism themselves and believe that AI is free of bias.

Analysing country clusters

In our analysis, we separate 120 countries into four clusters: low digital growth, emerging digital growth, advanced digital, and high innovation digital growth markets (Figure 8).

Figure 8: Digitally similar country clusters

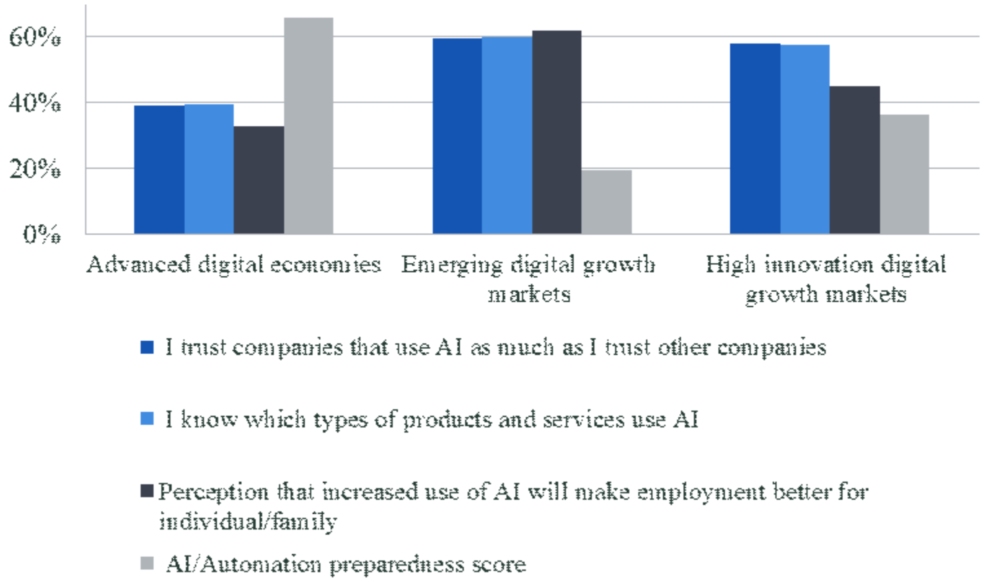


Source: SRI[44]

In January 2022, Ipsos released a country specific consumer survey covering investigating consumer sentiment for products and services using AI.[45] We mapped this data to our country clusters and found that once again, advanced digital economies with advanced AI/automated decision-making capabilities often lack consumer trust in perceived usefulness of AI (see Figure 9).

44 Digital trust II – A consumer perspective, Swiss Re Institute
45 Global opinions and expectations about artificial intelligence, Ipsos, 2022.

Figure 9: Country clusters and attitude/ perception of AI[46]



Source: SRI[47]

Analysis of the nine dimensions of digital trust provides insights into the factors and their relative quantitative importance in influencing degrees of digital trust. Some countries perform better on digital trust than others, without there necessarily being a correlation to presence and sophistication of digital infrastructure. There may be a gap between what institutions think is important to build digital trust (such as digital infrastructure, AI capabilities etc.) and what really matters to consumers (a sense of purpose, incentives, ease of use, and transparency of AI models).

CHAPTER 3: IMPORTANCE OF DIGITAL TRUST IN INSURANCE OUTCOMES

To truly harness the potential of new data-based technology, re/insurers must earn the trust of clients and consumers. According to the State of Digital Trust survey, 98% of financial services organizations recognize the importance of digital trust, but most believe that they have reached only accomplished a third of desired benchmarks.[48] Digital trust is very personal and highly emotional, not always responding to logical reasoning. Therefore, it can be hard to explain rationally why a digital insurance solution is trusted or not by customers.

In this section, we want to understand whether digital trust can contribute towards better insurance outcomes. The level of development of the insurance sector in a country can be gauged by its insurance penetration and density. Insurance penetration is calculated as the percentage of insurance premium to GDP while insurance density is calculated as the ratio of premium to population (per capita premium).

Several studies in existing literature look at factors impacting insurance density and penetration. Economic development of a country as measured by its GDP per capita is shown to play an important role in determining both insurance density and penetration.[49] This is an intuitive result of the fact that higher income results in both greater affordability and a higher level of wealth/possessions; hence higher demand for insurance products. Inflation, on

[46] Please note that IPSOS survey did not have data for countries we have categorized as low digital growth market; AI/ automation preparedness score has been developed as a part of our analysis, using data on different indictors- see Appendix 2

[47] Digital trust II – A consumer perspective, Swiss Re Institute

[48] State of Digital Trust Report, ISACA, 2023

[49] Oteng, Phyllis Asorh, Victor Curtis Lartey, and Amos Kwasi Amofa. “Modeling the Macroeconomic and Demographic Determinants of Life Insurance Demand in Ghana Using the Elastic Net Algorithm.” SAGE Open 13.3 (2023): 21582440231196658.

the other hand, negatively influences insurance consumption. Higher education is expected to be positively associated with insurance penetration and density since it results in better knowledge and awareness on risks. Another factor positively influencing insurance penetration and density is age dependency ratio - higher dependency ratio makes people want to buy more insurance cover to protect their dependents from financial hardships. Urbanization, unemployment and life expectancy are some of the other standard variables considered in statistical analyses of determinants of insurance density and penetration.

Our aim is to check if, after controlling for the impact of all standard macroeconomic variables affecting insurance penetration and density, we can find any evidence to suggest digital trust levels of countries positively impact insurance outcomes.

DATA AND METHODOLOGY

Our first task is to construct the variable of interest - digital trust scores. The only other study that has previously attempted to build an index to measure digital trust is the DII built by the Tufts University.[50] The DII is a combination of digital evolution and digital trust that considers four attributes - attitudes, behaviors, environment, and experience of the digital ecosystem.

We differ slightly in our approach towards constructing a digital trust score.[51] Our hypothesis is that digital infrastructure of a country and societal trust are both equally important in constructing a digital trust score. To that extent, we first compute a societal trust score, giving equal weightage to four parameters - trust in government, trust in businesses, trust in media and interpersonal trust. We believe that higher societal trust will translate into better digital trust, subject to the availability of digital infrastructure. Cross country data on trust in government and business is available from the Edelman Trust Barometer 2023, that surveys more than 32,000 respondents from 28 countries.[52] Data on trust in media has been taken from the Digital News Report 2023 published by the Reuters Institute.[53] Cross country data on interpersonal trust has been taken from the Ipsos survey on interpersonal trust across the world, that surveyed more than 22,000 respondents from more than 30 countries.[54] For countries with missing data, we have interpolated using regional averages.

We have then constructed the digital trust score using a geometric mean of digital availability and societal trust scores computed for 90 countries. This is our explanatory variable of interest. To study the impact of digital trust on insurance outcomes, we follow two approaches – ordinary least squares (OLS) regression for the cross section of countries and ensemble machine learning using random forest model. The former approach is used to analyze impact of digital trust on insurance density while the latter is used to study the impact of digital trust on insurance penetration.

Data on GDP per capita, insurance penetration, insurance density and inflation is from the Sigma database maintained by the SRI. Data on urbanization, age dependency ratio, adult literacy and internet penetration is from the World Bank while data on labor productivity is from the International Labor Organization . Data on digital development and national cybersecurity index have been taken from e-Governance Academy Foundation. The AI readiness index has been populated from Oxford Insights. The World Risks Report (2022) published by the World Economic Forum provides vulnerability score for different countries. The vulnerability score is composed of several socioeconomic, geopolitical, and demographic parameters that reflect a country's susceptibility as well as the lack of coping and adaptive capacities from shocks such as (but not limited to) natural disasters.

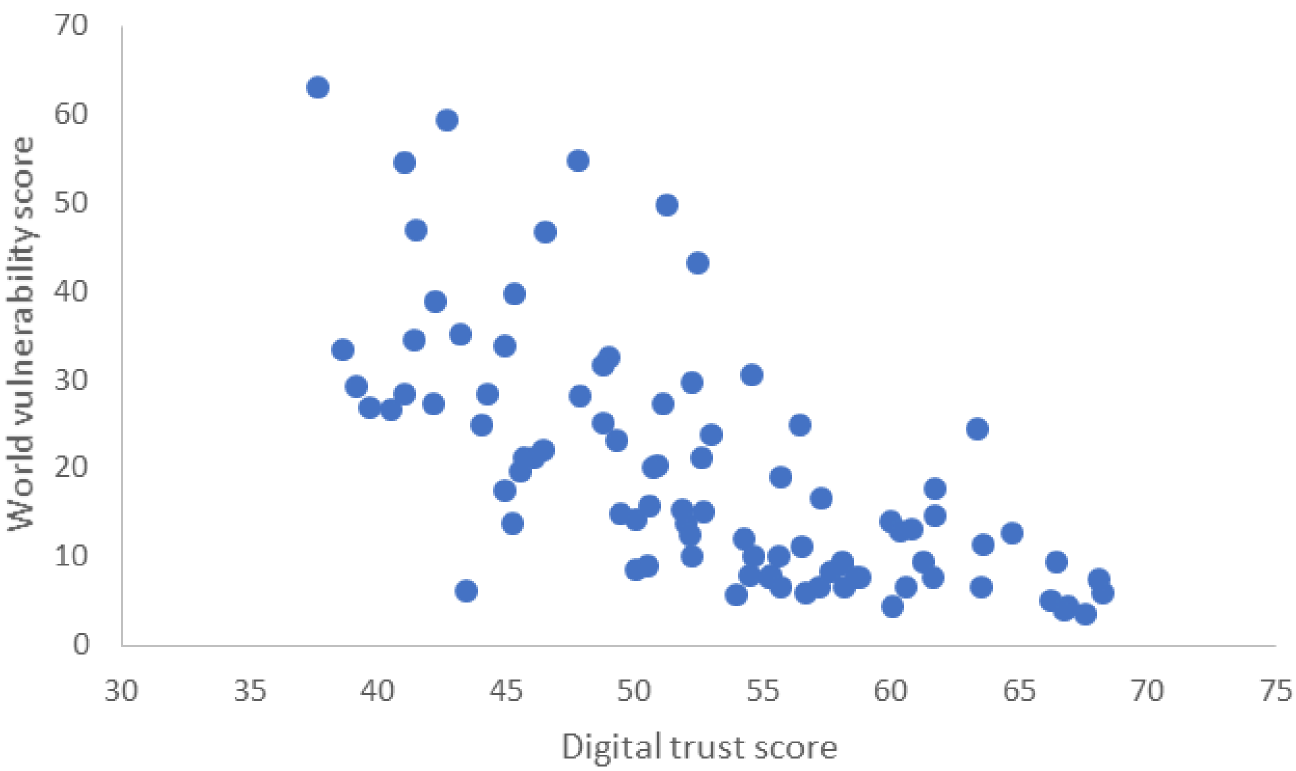
50 Digital in the time of COVID, Tufts University, 2019.
51 While the DII constructed by Tufts University is based on inputs to digital trust, in this section we focus on measurable outcomes of digital trust- hence trust as measured by government, businesses and society
52 Intelligence, Edelman. "2023 Edelman Trust Barometer." (2023).
Newman, Nic, et al. "Digital News Report 2023." (2023).
54 Interpersonal trust across the world, Ipsos, 2022.

RESULTS

OLS regression- Impact of digital trust on insurance density

We begin by testing if there is any association between vulnerability scores and our computed digital trust score. If our hypothesis is correct, we would expect countries with lower digital trust to be more vulnerable. When we draw a scatter plot between world vulnerability and digital trust it confirms our hunch on the association (Figure 10).

Figure 10: Scatter plot for digital trust and world vulnerability score



We then try to isolate the impact of all things digital from vulnerability scores (by regressing vulnerability on digital trust while also controlling for digital availability and artificial intelligence readiness) We use the residuals of this regression as a regressor for the next OLS cross-section that aims to estimate impact of digital trust on insurance density (after controlling for standard macroeconomic variables (per capita GDP, inflation, age dependency and the vulnerability residuals from the last step). The coefficient of digital trust turns out to be positive and statistically significant - a 1% improvement in digital trust scores can result in rise of insurance density by 76 points (i.e., a rise in insurance premium by 76 currency units, population remaining constant). The two equations and the statistical results are summarized below (Table 1).

$$VS_i = DT_i + \sum \alpha_i DC_i \ldots. (1)$$
$$\Delta VS_i = VS_i - \widehat{VS}_i \ldots. (2)$$
$$ID_i = DT_i + \sum \beta_i MC_i + \Delta VS_i \ldots. (3)$$

Where $VS_i, DT_i, DC_i, \widehat{VS}_i, \Delta VS_i, ID_i$ and MC_i stand for vulnerability score, digital trust score, digital control variables, predicted vulnerable score from (1), vulnerability residual from (2), insurance density and macroeconomic controls respectively.[55]

[55]While income levels, digital infrastructure, insurance penetration, education are all highly correlated factors within countries over time, only a weak correlation exists between countries

Table 1: OLS regression results[56]

Insurance density	Coefficient	Standard error	t value	P > | t |
Digital trust score	75.65	26.97	2.80	0.006
GDP per capita	.05	.01	7.78	0.000
Age dependency ratio	34.91	12.25	2.85	0.005
Vulnerability residuals	-10.86	8.03	-1.35	0.180
Constant	-6265.55	1614.84	-3.88	0.000

RANDOM FOREST MODEL- IMPACT OF DIGITAL TRUST ON INSURANCE PENETRATION

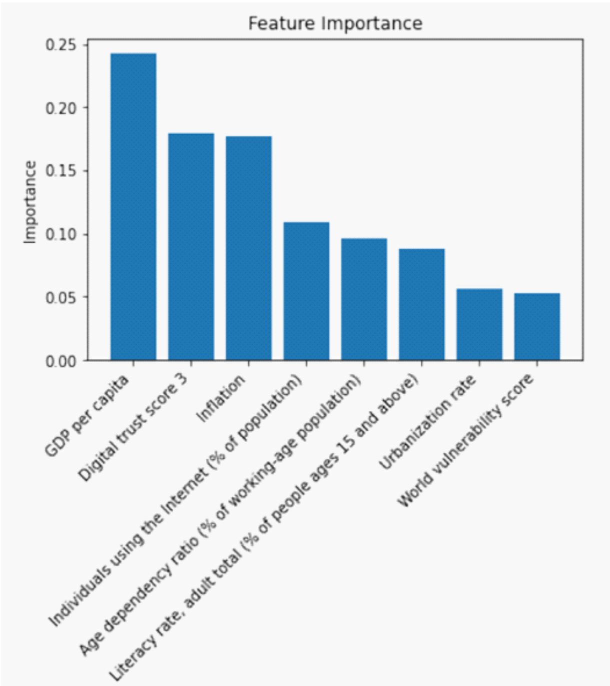
Ensemble learning is method in machine learning that reduces model instability by averaging predictions over multiple instances of similar models. Ensemble learning tends to reduce variance without significantly increasing bias, resulting in improved performance overall. We use a random forest model- that averages out predictions made by many individual decision tree models to get more accurate results. Each decision tree in the forest considers a random subset of features with access to a random set of data points. As with the previous exercise, we wish to check the importance of various factors in explaining insurance penetration in this case. The variables or factors under consideration include GDP per capita, digital trust score, inflation, unemployment, urbanization, age dependency, literacy rate, internet penetration, among others.[57] The feature importance of a random forest model depicts the relevance or importance of each variable X_i in the model based on the reduction in impurity, typically measured by the mean squared error, when using that feature for splitting across all nodes and trees:

Feature importance $(X_i) = \sum_j \frac{N_j}{N} \times \text{impurity reduction } (X_i, j)$

Where N is the total number of samples, Nj is the number of samples in node and impurity reduction (X_i,j) is the reduction in impurity achieved by using feature X_i to split node j.

As expected, in our random forest model, GDP per capita has the highest feature importance. Digital trust score has the second highest feature importance, showcasing its high importance in explaining insurance penetration. Figure 11 below summarizes this discussion.

Figure 11: Random forest model feature importance of different variables in explaining insurance penetration



[56] While we have tried various permutations and combinations of the regression analysis, we have reported results of the final regression exercise giving us the best goodness of fit

[57] Methodologically, randomization helps in reducing impact of highly correlated variables

To conclude this section, our results from both the statistical analysis (OLS and random forest model) help to establish our hypothesis that digital trust is an important contributing factor towards insurance outcomes - both in terms of density and penetration, across countries.[58] Therefore, insurers should actively seek to establish and strengthen digital trust with consumers.

CONCLUSION

For insurers, digitization changes how risk can be assessed and mitigated with new and more granular data. It enables improved underwriting in life insurance through more holistic and accurate risk pricing using electronic health data from wearables. Risk mitigation capabilities can also benefit, such as in the case of the automotive industry with Advanced Driver Assistance Systems. Similarly, SRI's resilience analysis suggests countries that are more digital are typically more resilient against other exposures, such a natural catastrophe risks.[59]

As with all other sectors, trust is relevant for the continued digitization of insurance business, when personal and other sensitive data are increasingly leveraged in different stages of the value chain, including in AI-training data. The notion of trust is largely subjective, making it difficult to measure. Surveys serve as the primary data source for quantifying consumer trust due to its elusive nature.

In this paper we first attempt to qualitatively list out nine factors that impact digital trust and divide these into three broad themes- reliability, security, and reassurance zone. We then quantitatively try to analyze which of these nine dimensions are most important in explaining digital trust. Our clustering exercise reveals that the three most important dimensions are cultural and generational attitudes, ease of use and explicability of AI/ automated decision making. We also divide countries into clusters- some countries perform better on digital trust than others, without there necessarily being a correlation to presence and sophistication of digital infrastructure.

This motivates us to construct a digital trust score, one which is not just dependent on the availability of digital infrastructure, but also on the societal notions of trust. Our analysis shows that countries that have greater digital trust scores have lower vulnerability to global risks. We also check for the importance of digital trust in explaining insurance outcomes- both insurance density and insurance penetration. While OLS results show that digital trust is highly significant in explaining digital trust, random forest model shows that digital trust constitutes the second most important variable (feature) in explaining insurance penetration across countries.

APPENDIX 1- DIGITAL MISINFORMATION AND ITS COSTS (CHAPTER 1)

I. Cost of digital misinformation on climate change

1. We consider 18 countries – United States, United Kingdom, Thailand, Switzerland, Spain, South Korea, Singapore, Russian Federation, Mexico, Japan, Italy, India, Germany, France, China, Canada, Brazil and Australia.
2. Population estimates (from World Bank) and social media penetration rates (from Statista) help us arrive at an estimate of total social media users across these countries.
3. Reuter's Institute Digital News Report 2021 gives us the share of adults who use social media as a source of news.
4. Survey research from Reuter's Institute[60] indicates that about 9% of respondents refer to social media for news related to climate change.
5. SRI analysis, based on secondary literature, estimates that approximately 20% of climate change-related news on social media is driven by misinformation.
6. For the purpose of this analysis, based on secondary literature, we peg the probability of getting influenced by climate change-related fake news at 30%.

[58]We specifically use OLS for insurance density since it does not have dependency on income levels. Randomization is better as handling inter variable dependencies- hence insurance penetration which could be highly correlated with income (by definition) is modelled using random forest.
[59] The economics of digitalization in insurance., Swiss Re, 2023.
[60] S. Andi, "How People Access News about Climate Change", Reuter's Institute, 2021.

7. The multiplicative impact of steps 4, 5 and 6 on total social media users helps us arrive at an estimated number of people persuaded to climate inaction due to fake news, for each country.
8. Studies suggest that the economic cost of climate change per person (accounting for both cost to humanity as well as impact of climate change on economic growth) could be as high as USD15,000.[61]
9. For each country in our analysis, we multiply the number of people driven to climate inaction with the per person economic cost of climate change (adjusted for purchasing power parity) to arrive at a dollar value of climate inaction.
10. Since the countries in our analysis comprise 85% of the global GDP, the sum total of the cost of climate inaction across the 18 countries can be extrapolated to arrive at a global cost of digital misinformation-driven climate inaction - USD 37.5 billion, which is about 0.05% of global GDP.

II. Cost of digital misinformation on non-communicable diseases

1. We consider the same set of 18 countries as before.
2. Population estimates for the age group 30-70 (from World Bank) and social media penetration rates (from Statista) help us to arrive at an estimate of total social media users in the target age group across these countries.
3. Based on secondary literature and our own understanding, we make the following assumptions for this analysis: namely, that 60% of the population uses social media for information on diet and exercise; that 40% of the social media content on these topics is driven by misinformation; and that the probability of influence is 30%.
4. The multiplicative impact of assumptions in step 3 helps us to arrive at an estimated number of people influenced by fake news on diet and exercise, in each country.
5. We factor in the NCD YLLs for each country from the Global Burden of Diseases database and, using the assumptions in the previous step, arrive at the proportion of NCD DALYs due to misinformation.
6. We assume per capita income of a country to be a proxy for the loss associated with each DALY. Multiplying the total NCD DALYs driven by digital misinformation with the per capita income provides an estimated opportunity cost (premature mortality) for digital misinformation driven NCDs for a country.
7. Since the countries in our analysis comprise 85% of the global GDP, the sum total of the economic cost of NCD driven DALYs across the 18 countries can be extrapolated to arrive at a global cost of NCDs driven by digital misinformation - USD 87.4 trillion, which is about 1.8% of global GDP.

APPENDIX 2: QUANTITATIVE ANALYSIS OF DIGITAL TRUST PYRAMID (CHAPTER 2)

I. **Identification and grouping of parameters under digital trust dimensions:** SRI study on digital trust last year scopes out the prerequisites to trust digital trust pyramid.

Our literature survey has found an extensive list of parameters relevant to each of the nine digital trust dimensions. The parameters were initially filtered based on the robustness of data methodology, geographic coverage, and reliability of source. Furthermore, we used Cronbach's alpha measure of reliability to evaluate the grouping/fit of parameters between the digital trust dimensions. Reliability of any given parameter here refers to the extent to which it is a consistent representation of a digital trust dimension, and Cronbach's alpha is one way of measuring the strength of that consistency.

Cronbach's α is:

$$\alpha = \frac{\mathrm{N}^2\overline{\mathrm{Cov}}}{\sum \mathrm{S}^2_{\mathrm{item}} + \mathrm{Cov}_{\mathrm{item}}}$$

The top half of the equation is simply the number of items (N) squared multiplied by the average covariance between items. The bottom half is just the sum of all the item variances and item covariances.

Through this process, we were able narrow our analysis down to the most suitable signals.

[61] Economic cost of climate change could be six times higher than previously thought, UCL News, 6 September 2021.

Parameters considered for the assessment

Digital trust dimension	Parameters	Source
Access to internet/data	Digital development level	NCSI, e-Governance Academy Foundation, 2021.
	Individuals using the Internet (% of population)	ICT indicators database, ITU, 2022.
	Fixed broadband subscriptions (per 100 people)	ICT indicators database, ITU, 2021.
	Population covered by at least a 3G/4G mobile network (%)	ICT indicators database, ITU, 2020.
	Households with internet access at home (%)	ICT indicators database, ITU, 2021.
Ease of use/accessibility	Households with a computer at home (%)	ICT indicators database, ITU, 2021.
	E-government development	UN E-government knowledge database, UN, 2022.
	Made or received a digital payment, female (% age 15+)	The Global Findex Database, World Bank, 2021.
	Used a mobile phone or the internet to pay bills, female (% age 15+)	The Global Findex Database, World Bank, 2021.
	Technology readiness score	Portulans Institute, 2022.
Cultural/generational attitudes	Regulatory quality	Worldwide Governance Indicators Project, World Bank, 2021.
	Rule of law	Worldwide Governance Indicators Project, World Bank, 2021.
	Govt effectiveness	Worldwide Governance Indicators Project, World Bank, 2021.
	Governance score	Portulans Institute, 2022.
	Corruption perception	Transparency International, 2021.
Data/cyber risk	Global cybersecurity score	ITU, 2021.
	National cyber security	e-Governance Academy, 2021.
	Secure internet servers (per 1 million people)	Secure internet servers, Netcraft and World Bank, 2020.

Data ethics/privacy	Privacy protection by law content score	Portulans Institute, 2022.
	Governance and ethics score	Government AI readiness, Oxford Insights, 2022.
	Legal framework's adaptability to digital business models	World Economic Forum, 2019.
	Regulations, policies, and guidance provide a comprehensive framework for generating and publishing open data (extent)	Global Data Barometer, 2021.
Security by proxy: laws, kitemarks	E-participation	UN E-government knowledgebase, United Nations, 2022.
	ICT regulatory tracker score	ICT regulatory tracker, ITU, 2022.
	Relevant laws, regulations, policies, and guidance provide a comprehensive framework for protection of personal data (extent)	Global Data Barometer, 2021.
AI & automated decision making	AI readiness	Government AI readiness Index, Oxford Insights, 2022.
	Global innovation score	Global Innovation Index, WIPO, 2022.
	PCT patent applications	Portulans Institute, 2022.
	Research and development expenditure (% of GDP)	UNESCO Institute for Statistics, 2022.
Empathetic/social AI	Chatbot search interest[62]	Google Trends, Google, 2022.
	Data representativeness score	Government AI readiness Index, Oxford Insights, 2022.
	Scientific and technical journal articles	National Science Foundation, Science and Engineering Indicators, World Bank, 2021.
Human interaction	Multi-stakeholder collaboration	Global Competitiveness Index, WEF, 2019.
	Competition in services	Global Competitiveness Index, WEF, 2019.
	Firm-level technology absorption	Global Information Technology Report, WEF.

[62] Some chatbots based on LLM that have the ability to constantly fine tune basis prompts could be proxies for cognitive engagement- in this context chatbot interest allows us to measure the demand and scope for interaction between humans and machines. Our expectations are that the quality of chatbot responses will improve so that in many use case scenarios, users may not be able to differentiate between chatbot and human, thus enhancing the aggregate perception of empathy among customers.

ii. Clustering and analysis of variance: We then try to analyze the combination of indicators that have the highest power to explain/ contribute to digital trust. We use K-means cluster analysis, ANOVA and principal component analysis (PCA) for this quantitative analysis.

Initial variance analysis reveals that cultural and generational attitudes, and ease of digital use, and explainability of AI/ automated decision making to be the biggest contributors to digital trust.

Calinski-Harabasz criteria were used for identifying the optimal number of clusters and for assessing the cluster quality. The criteria are defined as,

$$\frac{SS_B}{SS_W} \times \frac{N-k}{k-1}$$

where $SS_B$is the overall between-cluster variance, SS_W the overall within-cluster variance, k the number of clusters, and N the number of observations. The greater the value of this ratio, the more cohesive the clusters (low within-cluster variance) and the more distinct/separate the individual clusters (high between-cluster variance).

Analysis of variance (ANOVA) is a collection of statistical models useful for analyzing variation within and between observations that have been partitioned into groups or clusters. Analysis of variance was computed per variable, and the resulting analysis of variance table was used to determine which variables are most effective for distinguishing clusters.[63]

ANOVA statistics for K-means clustering:

Dimension	SSB	SSW	F-statistic	p value
Cultural /generational attitudes	95.8	23.1	241.6	0.000
Ease of use	94.3	24.6	223.5	0.000
AI explainability	92.5	26.4	204.7	0.000
Access to data/access to the internet	85.4	33.6	148.4	0.000
Data ethics/privacy	83.7	35.2	138.8	0.000
Empathetic/social AI	77.8	41.1	110.5	0.000
Trust by proxy: Laws, self-policing	74.4	44.5	97.6	0.000
Data/cyber risk	73.9	45	96	0.000
Human to human interactions	57.3	61.6	54.4	0.000

[63] Please note that we are only trying to ascertain the relative importance of dimensions associated with digital trust based on separation of clusters; we are not establishing causation

GCA 24

DATA, DISRUPTIONS AND THE ACTUARY

12, 13, 14 FEBRUARY, 2024 AT THE WESTIN POWAI LAKE, MUMBAI

PAPER III

An investigation into potential ways to bring the old and new Indian government pensions closer together

An investigation into potential ways to bring the old and new Indian government pensions closer together

Kulin Patel, FIAI, FIA
Ganesh Sudrik

ABSTRACT

The Government of India decided to embark on a defined contribution pension path (with the introduction of the National Pension System - NPS) for its own employees, followed by State Government and public sector companies, over the last 20 years. However, a potentially dangerous trend in the narrative has begun to take shape in recent times with certain States deciding to revert to the Old Pension Scheme (OPS). There are reported hunger strikes on the topic as recently as 9 January 2024.

The decision to revert to OPS, by some States, seems to have been motivated by short-term considerations, as well as pressure from employees citing the retirement income from NPS would not be comparable to the OPS. Although, we would hope, that stakeholders do understand the issue with OPS' associated long term risks, and potential impact on public finances. These long-term risks are the same as with any large unfunded defined benefit scheme. Most typically, uncertainty of long-term costs, increasing longevity and the difficult task to balance public finances.

This paper aims to investigate the extent to which a couple of simple interventions could possibly help bridge the pension income differences between OPS and NPS. We then introduce the reader to a conceptual post-retirement income longevity risk sharing solution that could further be used to increase potential retirement incomes. However, no quantitative analysis for longevity pooling is included in this paper. The topic could be taken up in further studies.

Keywords

old pension scheme, national pension system, income drawdown, longevity pooling, investment allocation

Correspondence details

*Kulin Patel, CEO/ Partner, K.A. Pandit, C201-Remi Bizcourt, Off Veera Desai Road, Andheri (West), Mumbai 400058.

Email: kulin@ka-pandit.com

INTRODUCTION AND BACKGROUND

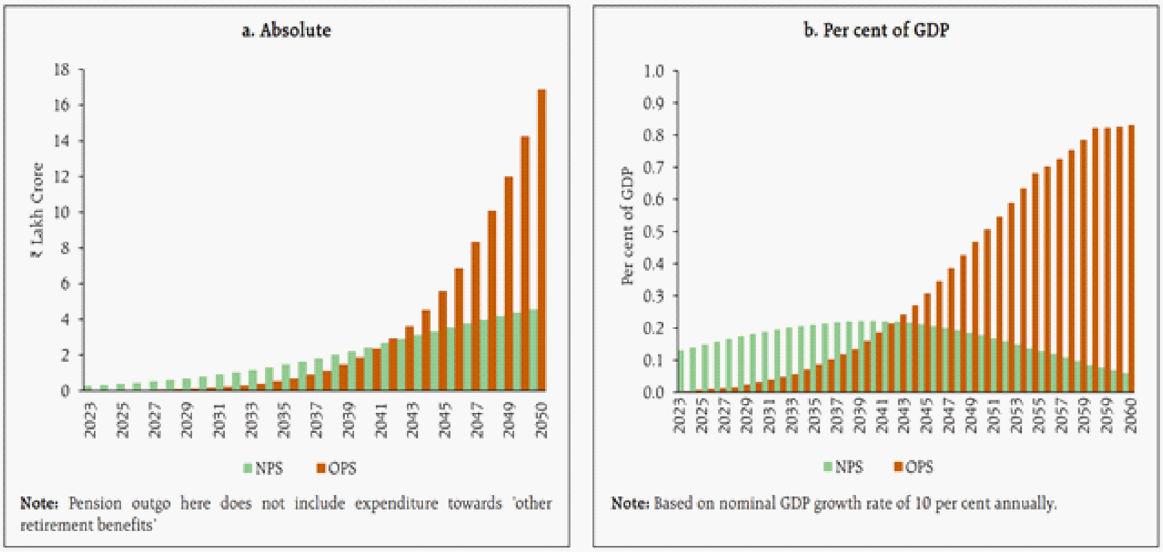
Significant work was done in the early part of this century culminating in a major decision, by the Government of India, to reform the escalating unfunded pension liabilities associated the defined benefit pension schemes of Central and State government employees. The government introduced the National Pension System in 2004, a defined contribution scheme designed to replace the OPS, which was a DB plan. This transition was not only a national initiative but also extended to most State governments in India **(See Appendix 1)**. As of 31 October 2023, there were just under 25 lakh (2.5 million) Central and over 63 lakh (6.3 million) State Government employees in the NPS.[1] In total, this represents over 50% of the entire core NPS subscriber base (excl. Atal Pension Yojana)[1]

A detailed article in the archives of the Reserve Bank of India published in September 2023 "Fiscal Cost of Reverting to the Old Pension Scheme by the Indian States – An Assessment"[2] importantly projects the financial impact on State finances if the State NPS employees reverted to OPS. The article has the full details of the methodology and assumptions the authors had used.

We show a few extracts from that report as context of our paper. One important fact to keep in mind when contextualising each graph is that the NPS outgo figures are the projected contributions paid by the governments into an active employee's defined contribution NPS account. However, for the OPS, the outgo is the actual pension payable to the retiree population.

The projected outgo graph clearly shows how the OPS outgo (in dark orange) would remain low for several years as these NPS employees will not have reached retirement. However, in about 20 years once they start retiring, the OPS outgo rapidly increases. As a percentage of GDP the gap the difference in outgo, between the NPS and OPS, by in 25 years' time could be as wide as 2x. In about 35 years that gap may widen to about 7x (c. 0.1% of GDP in NPS versus c. 0.7%).

Chart 6: States' Projected Pension Outgo – NPS *vis-à-vis* OPS



Note: Pension outgo here does not include expenditure towards 'other retirement benefits'

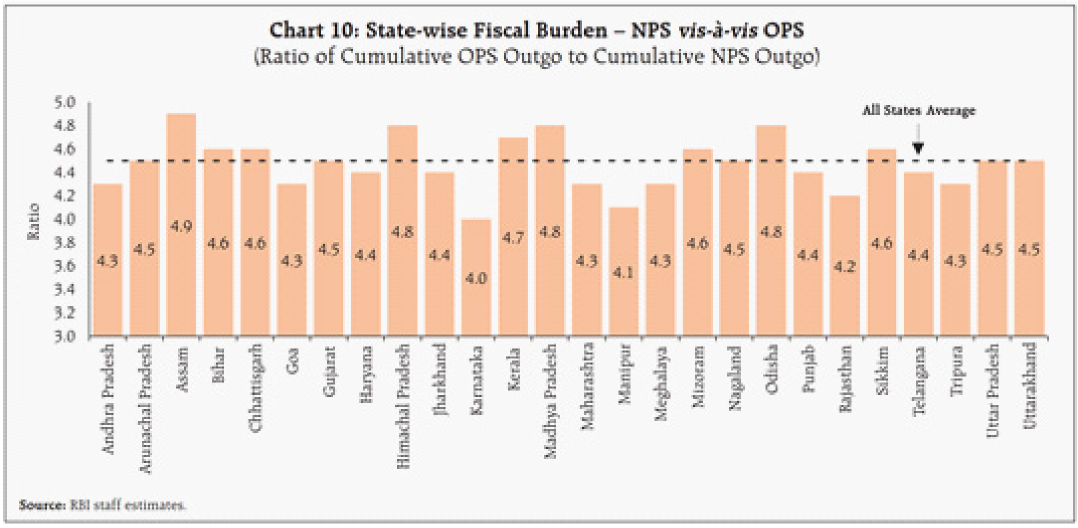
Note: Based on nominal GDP growth rate of 10 per cent annually.

Source: RBI staff estimates.

The next graph really highlights the difference in the outgo. The additional cost to the States after about 18 years is so high. Given the NPS outgo (contributions) will be higher than OPS outgo in the initial years, one can see precisely why decision makers may have been tempted to revert to OPS. Even more so, given the employee sentiments.

1 Pension Regulatory and Development Authority Pension Bulletin October 2023 https://www.pfrda.org.in/myauth/admin/showimg.cshtml?ID=2795
2 Fiscal Cost of Reverting to the Old Pension Scheme by the Indian States – An Assessment by Rachit Solanki, Somnath Sharma, R. K. Sinha, Samir Ranjan Behera and Atri Mukherjee https://rbidocs.rbi.org.in/rdocs/Bulletin/PDFs/02AR18092023A646C561E9BC438AA67BFD934A58672E.PDF

Taking the single cohort of State employees as of March 2022 (no new entrants assumed), the article outlines that the ratio of present value of total OPS burden to the present value of total NPS burden, over the period end-March 2023 to end-March 2084, will increase on an average by around 4.5 times if States choose to shift from NPS to OPS.



The actual situation can be even more acute given that the RBI considered a uniform expected lifetime to age 80.[2] By 2060, India is projected to have an expected lifetime age of 83.5 once you have reached age 60.[3]

OUTLINE OF OUR STUDY AND APPROACH

One of the main narratives in the OPS vs NPS issue has been around the fact NPS cannot generate income levels for the State employees to the extent of the OPS. A secondary angle is that the NPS contribution rate is not enough to replicate the OPS level of benefits.

What if there are simple ways to increase the chances of generating higher income from NPS, without additional contributions? Furthermore, what if we expressed the actuarial cost of OPS in a different way, in case it helps provide a different angle to the narrative.

We want to estimate the how far asset allocation changes and, a potential of a lifetime income drawdown approach can go towards bridging the perceived mathematical gap of the NPS and OPS outcomes for retirees.

As much as the specific numbers/ absolute values in our paper, our focus is to introduce the concepts. We recognise there are many rigourous techniques, approaches and assumptions that could have been used. We have tried to balance simplicity and rigour as a starting point for the purposes of this paper. We have not attempted to comment on the practical, regulatory and policy level considerations, given the purpose of this paper to be included in an actuarial journal for the Global Conference of Actuaries. We have therefore also not fully considered all current regulatory limitations or designs in place today. We hope our work will be a starting point for further deeper studies into the topic as it applies to India.

Whilst our motivation was from the State government decisions, we believe much of the work in this paper could be easily adapted and applied to any of the public sector companies, public sector banks and of course Central Government.

Below is a summary of the basic approach and methodology used for each part of our investigation. In all cases, we used a cashflow model, projecting benefits, accumulated balances etc. for each year into the future for sample lives.

[2] Fiscal Cost of Reverting to the Old Pension Scheme by the Indian States – An Assessment by Rachit Solanki, Somnath Sharma, R. K. Sinha, Samir Ranjan Behera and Atri Mukherjee https://rbidocs.rbi.org.in/rdocs/Bulletin/PDFs/02AR18092023A646C561E9BC438AA67BFD934A58672E.PDF
[3] United Nations, Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022, Online Edition.

We have used nominal values across the board and have not considered any taxation effects.

There are features of our approach, assumptions or modelling that could be enhanced in future studies. Some approximations have been made for ease of computation and speed. Some thoughts on where enhancements might be made in future studies are outlined in the closing summary section.

The quantitative portion of our paper builds up a story in three (3) parts.

For a key summary of the OPS benefits and a comparison with NPS, see **Appendix 2**.

Part #1

If an employee joined OPS today, then what would be an estimated standard contribution rate, using an entry age methodology and how does that compare to the current contribution into NPS (we take 24% of basic salary plus DA).

More specifically, with our assumptions, we demonstrate the confidence interval within which the NPS contribution rate fits into the entry age OPS estimated contribution rates.

Standard Contribution Rate

We have used an entry age method to estimate the standard contribution rates for the OPS at each entry age. We have conducted the calculations for entry ages 20-35.

The standard contribution rate under this method is determined as the contribution rate which, if payable over the expected future membership of a group of new entrants, would provide for the total expected benefits payable in respect of that group.

For a specific life it is the:

Present value of total projected benefits ÷ Present value of projected future salaries

Benefits projected include payment by way of expected pension, widow/er and family pension, additional quantum pension after attaining milestone ages, withdrawal benefits for actives and pension in payment including widow/er pension in payment etc.

The twist we added is that the discount rate is a random generated set of interest rates (see returns data and assumption section). These interest rates will also be used for the baseline accumulation of the NPS contributions under Part #2. For each year we ran 1,000 scenarios.

The result is a range of entry age contribution rates for each modelled entry age.

Part #2

Compare projected retirement income streams from age 60 under OPS and NPS along with the current default NPS annuity and fund management approach for the State employees.

How does that comparison change if we generated a range of potential NPS income streams by altering the pre-retirement asset allocation to the Lifecycle auto choice allocation?

For some sample lives, we plotted the projected retirement pension from age 60. To illustrate the concept, we have plotted the employee's retirement pension only. This means, only the single life stream, as well as, only showing the post lump sum/ commutation pension.

For the OPS, the pension is determined by the defined benefit formula. We have included dearness relief increases in payment as per the OPS.

NPS – Accumulation Default Asset Allocation

For the NPS we firstly arrived at an accumulated balance at retirement age. We have initially accumulated the contributions in line with the same interest rates as used for the OPS discount rates as in part #1. As the OPS is "pay as you go", one could argue the discount rates should be the cost of capital the State's can raise money. On the other side, the default asset allocation for the NPS funds managed for the Government sector is mostly in central government, state government bonds, and other fixed interest. Although there are differences, for the broader concept of our study, it does not seem unreasonable to use the same rates. It also ensures some level of consistency as we want to compare OPS with NPS. (see assumption section)

NPS – Accumulation Alternative Asset Allocation

For the second aspect of this part #2 we changed the asset allocation to generate a set of annual returns per the prescribed NPS Lifecycle Auto-Choice norms (LC). In particular, the LC50

(Lifecycle with 50% in Equity up to age 35) and LC25 (Lifecycle with 25% in Equity up to age 35) asset allocations (see assumption section). The auto-choice has been an option for Central Government employees since 2019.

NPS – Annuity Conversion (see Appendix 3)

For NPS, we needed to convert the accumulated balances into a pension. Currently, in practice, the balances are converted using the annuity service providers in the overall NPS architecture. We obtained the average annuity rate available for Government NPS subscribers from the Protean (NSDL) website. There was a large variation in the rates between providers. Hence, we used a simple average. A big assumption is we have not adjusted the annuity rates for future anticipated changes in market conditions. i.e. the rate derived as on 21 December 2023 are assumed to apply in the future, or one can say we are assuming retirement is on 21 December 2023.

Another issue is that NPS has many options for the form of the annuity the subscriber wishes to receive as a benefit. There is no annuity form available under NPS that matches the spouse's pension of the OPS. For the purposes of converting the NPS balance to an income, we have used the NPS market annuity rates to derive an annuity that matches the OPS 60% spouse's pension benefit. The precise approach is illustrated in Appendix 3. We have used the rates without return of purchase price, so it is a closer comparability to OPS.

One other benefit where NPS annuities available do not match the OPS benefit is the pension escalation. i.e. there are no annuities available under NPS that give indexed increases in payment. Despite this, we did not try to adjust NPS annuity rates for this point.

We assumed there would be no other costs associated with the annuity (i.e. the annuity rate incorporates all costs).

Part #3

How does the comparison change if we generated a range of potential NPS income streams by altering the pre-retirement asset allocation was changed to the Lifecycle auto choice allocation, and the post-retirement income was from a lifetime income drawdown approach.

We now take Part #2 to another level. The NPS has recently introduced a systemic withdrawal option for a portion of a subscriber's balance at retirement.[4] The new option is only for a portion of the corpus and only until the age of 75. However, for the concept in our study where we are comparing with OPS, we assume the balance is available for a lifetime income drawdown option.

In terms of asset allocation, we move the current asset allocation under the lifecycle auto-choice to start at age 50 rather than the current 35. This is to reflect a longer investment horizon now we are to introduce income drawdown.

We have used the following approaches for other aspects of the income drawdown:

- The starting balance is taken from the median balances generated from the pre-retirement accumulation.
- The income withdrawal each year is set at a level pension that will aim to exhaust the balance by the expected lifetime from the age in that year. The underlying mortality table is stated in the assumption section. A more sophisticated version of this is referred to in a Pension Decumulation Pathways Working Party Paper of the Institute and Faculty of Actuaries published in May 2022,[5] as the notional annuitisation approach. Our difference being we did not allow for future investment rates in the self annuitisation calculation. We used the expected lifetime as a straight-line calculation on the balance at each age.
- Each year, the withdrawal is assumed to be taken uniformly through the year, and we applied the estimated generated investment return based on a weighted average return of the asset allocation in the adjusted lifecycle strategy.
- The above is repeated each year and we plotted the resulting annual income stream.

There are many alternative approaches for

[4] PFRDA Circular PFRDA/2023/30/SUP-CRA/10 dated 27 October 2023

[5] Pension Decumulation Pathways A proposed approach Pension Decumulation Pathways Working Party, May 2022 – Institute and Faculty of Actuaries

income drawdown paths, and this would be a good topic for further studies within the actuarial profession in India. We hope our investigation, however, helps as a start, as well as, illustrating the concept.

DATA USED FOR MODELLING

Employees data

We conducted the analysis on sample lives (model points) using a cross section of characteristics. The following sample lives were used in terms of age at entry, salary level.

Age at entry: 25-35 years

Salary at entry: 50,000 per month basic salary + 46% as Dearness Allowance

Past Investment Returns

We obtained daily Net Asset Values for every fund manager, each of the asset category from the PFRDA Handbook[6] since the inception of each manager up to 31 October 2023.

State Government Returns

In particular, the three fund managers used for the State Government (SG) NPS, namely SBI, LIC and UTI have a specific NAV for the State Government funds (SG) they manage. The first date available for the State Government NAV was 27 February 2009.

A log normal daily return was calculated for each fund manager. An overall average daily return was calculated, under the understanding that the Government has been giving each fund manager an equal share of the contributions to invest. An annualised yield was derived as the average log normal daily return, multiplied by the average number of trading days in a year.

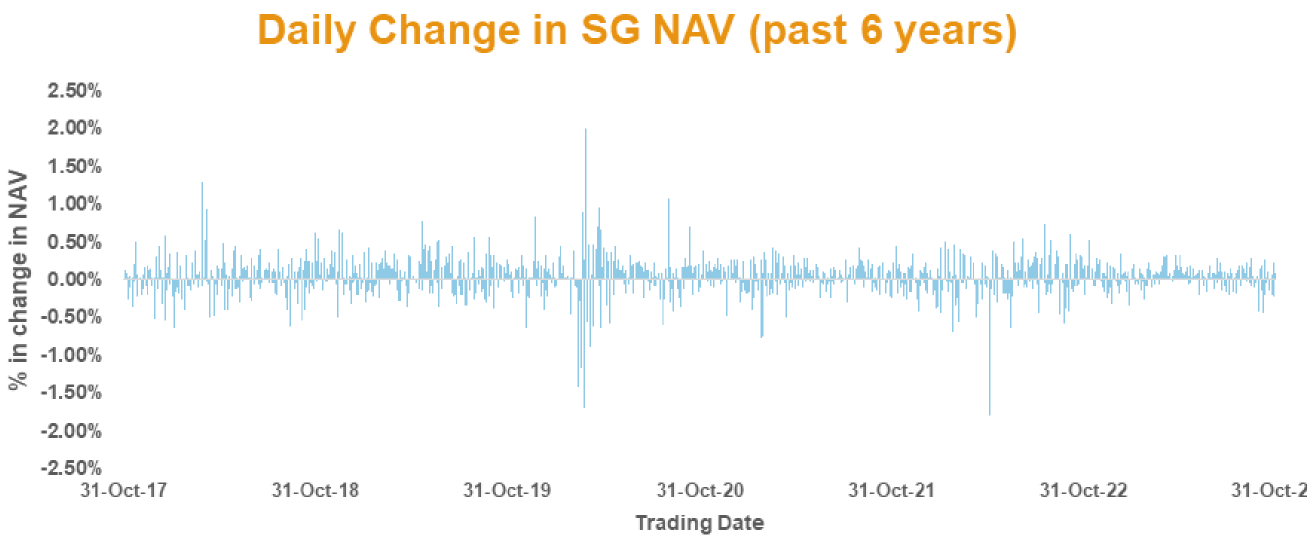
Below is a summary of the calculated mean annualised returns and standard deviations.

Graph for the change in the last 6 years' daily returns is shown below.

% p.a.	Since inception	10 years	6 years	5 years	3 years
Mean	8.49	9.01	7.45	8.72	6.42
Std. Dev.	3.66	3.48	3.63	3.61	3.00

Table 1: SG NAV Returns

Graph 1: Daily change in SG NAV



Private Fund Managers

As an input for our alternative investment scenarios for Part #2 and #3 of our study we have used a similar approach as above but for each private fund managers and each Tier I[7] asset category (as defined by PFRDA for the NPS).

Below is a summary of the calculated mean annualised returns and standard deviations.

E (Equity)

% p.a.	Since inception	10 years	6 years	5 years	3 years
Mean	10.98	12.08	10.59	13.43	17.78
Std. Dev.	15.27	15.88	17.58	18.65	14.03

Table 2: Tier 1 E - NAV Returns

C (Corporate Bonds)

% p.a.	Since inception	10 years	6 years	5 years	3 years
Mean	8.82	8.58	7.03	7.98	5.07
Std. Dev.	2.77	2.75	2.91	2.97	1.80

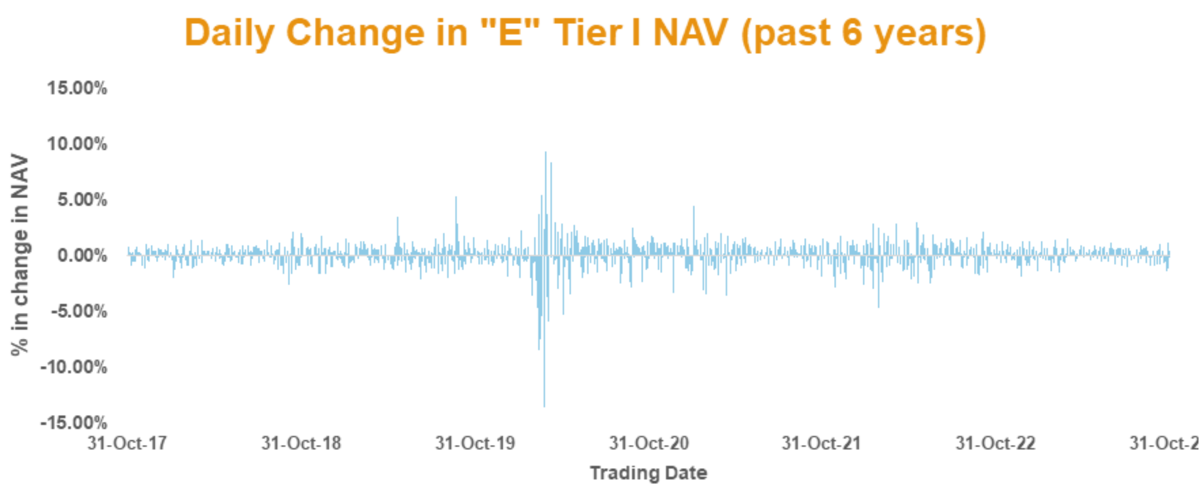
Table 3: Tier 1 C - NAV Returns

G (Government Bonds)

% p.a.	Since inception	10 years	6 years	5 years	3 years
Mean	8.03	8.65	7.06	8.16	4.14
Std. Dev.	4.01	3.92	3.85	3.70	2.83

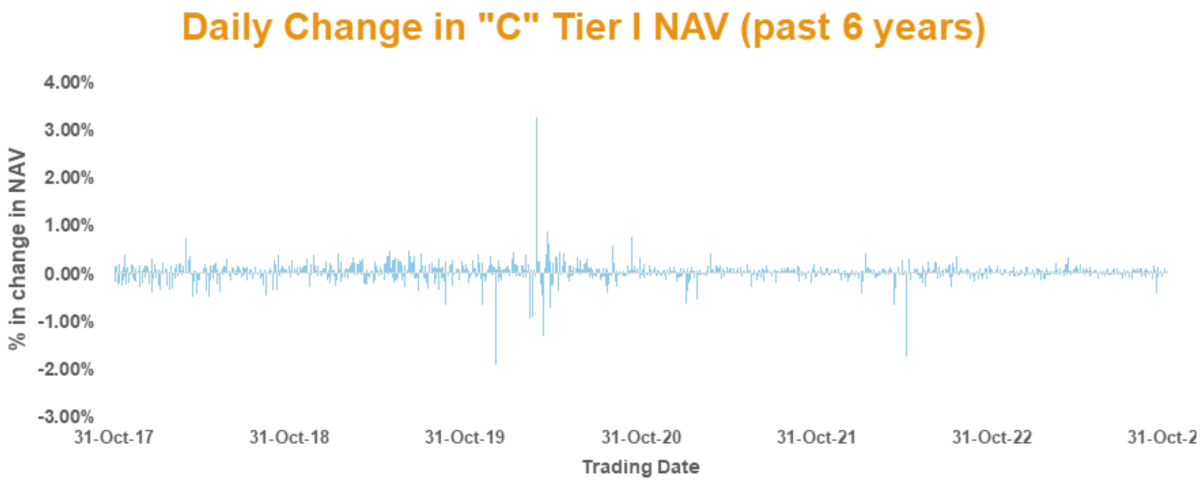
Table 4: Tier 1 G - NAV Returns

Graphs for the change in the last 6 years' daily returns are shown for illustration below.

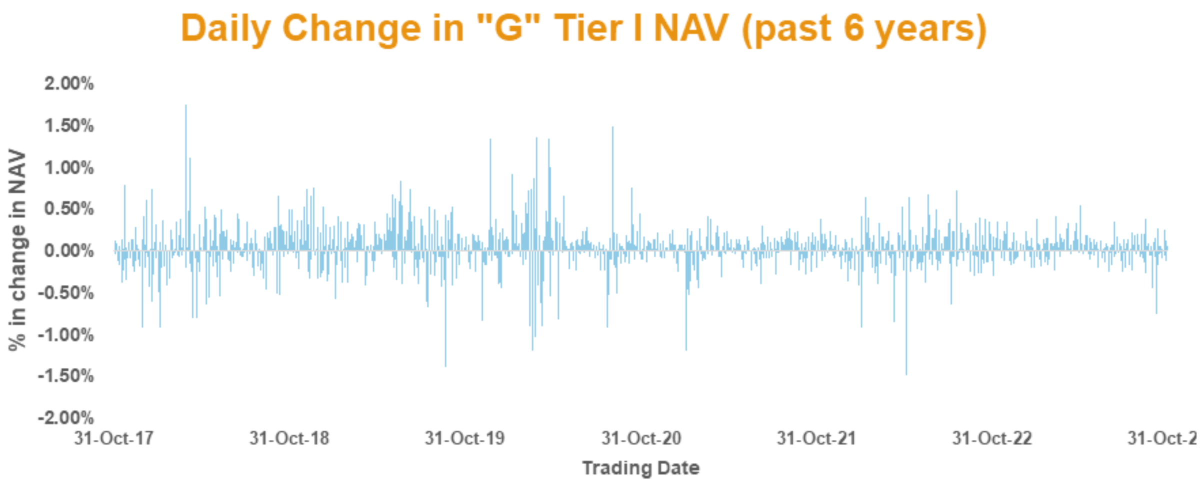


Graph 2: Daily change in E Tier 1 NAV

[7] We focussed only on NPS Tier I NAVs



Graph 3: Daily change in C Tier 1 NAV



Graph 4: Daily change in G Tier 1 NAV

ASSUMPTIONS - FINANCIAL ASSUMPTIONS

Discount rates: In order to ensure a like for like comparison for OPS and NPS, as well as considering that OPS is funded itself through the Government, we have **used a mean of 7.5% p.a. and standard deviation of 3.5% p.a.** to generate random future returns for the State Government managed funds, using a normal distribution. We generated 1,000 investment return path scenarios. There is an inherent assumption that the current default asset allocation remains unchanged in the future for the OPS/ NPS.

We recognise past performance will not be the best indication of future rates and one should use a suitable economic scenario generator. Given the returns are going to be used both for the OPS discount rate and NPS accumulations, we are comfortable with the approach as we will still get to see the comparisons.

Expected Investment return: We have used the following mean annualised returns and standard deviation for the NPS accumulations to similarly generate 1000 scenarios for each year in the future, using a normal distribution.

% p.a.	Mean	Std. Dev.
SG Default	7.5	3.5
E – Tier I Private Sector	12	15
C – Tier I Private Sector	8	2.75
G – Tier I Private Sector	7	3.5

Table 5: NAV Returns Normal Distribution assumption

Salary Growth: We have used a deterministic assumption 12% p.a. as used in the RBI article. This includes increments due to merit, salary revisions and dearness allowances, with half of the contribution coming from dearness allowance. This overall 12% CAGR works out to be in the range of what we have seen from our experience on public sector pensions.

Pension increases in retirement: As used in the RBI article, we have used a pension increase rate of 6% p.a. This is half the salary growth rate and is also within the range of the RBI long term inflationary range.

OTHER ASSUMPTIONS

Post Retirement Mortality: For Part #1, and where the NPS accumulations are modelled for the income drawdown, we have used the latest published annuitants table, available on the Institute of Actuaries website as of 15 December 2023. Namely the Indian Individual Annuitant's Mortality Table (2012-15).[8] Given we are comparing OPS and NPS and therefore the specific values of our findings are less important, we have not adjusted the table for future improvements in mortality.

Pre-Retirement mortality: Given the focus of our study is largely on the post retirement income phase, and the fact that modelling defined contribution projections using sample lives gets distorted because we will not consider separate costs of death benefit – we have assumed no pre-retirement mortality in any of our modelling. We realise this distorts the estimated cost of the OPS; however, it will mean a greater comparability with NPS.

Attrition: Similarly, we have not assumed any pre-retirement attrition. We know however that attrition in the Government sector is very low.

Lifecycle Asset Allocation

For Part #2's alternative asset allocation we have used the LC50 and LC25 asset allocation that are prescribed by PFRDA in the current NPS architecture. Sample allocations are shown below. A full table is found at www.pfrda.org.in

	LC50			LC25		
Age	E	C	G	E	C	G
Upto 35	50%	30%	20%	25%	45%	30%
40	40%	25%	35%	20%	35%	45%
45	30%	20%	50%	15%	25%	60%
50	20%	15%	65%	10%	15%	75%
55	10%	10%	80%	5%	5%	90%

For Part #3, we have moved the above percentages to start age 50 instead of 35. This is to reflect a longer investment horizon when we introduce income drawdown.

Form of income stream in retirement: For Part #1, we have considered all potential future cashflows in retirement to work out the full actuarial value (including retiree and spouse' pension). However, when we look at Part #2 and #3, we focus on the income stream for the original retiree only. We believe this should be suitable enough to highlight the concepts we want to introduce, without the introducing the complexity.

Special Note on sensitivity of salary growth assumption

The rest of the paper builds upon a narrative based on a 12% salary growth assumption. Of course a lower rate could have been used which may alter the narrative somewhat. However, actual past experience, and an objective source for the salary growth assumption of 12% is a key reason we retained it for the study. Another reason is that public sector pay has been much higher than inflation, investment returns in the past. If the main study was to use a 8% salary growth, then we would need to reconsider the calibration of our expected return assumptions too. This may negate the effect of reducing the salary growth assumption. However, as an indication, we have included **Appendix 4** that

8 https://actuariesindia.org/sites/default/files/2022-05/Indian_Individual_Annuitants_Mortality_Table(2012_15).pdf

gives a short summary of some results by only changing the salary growth to 8% p.a. No other assumptions have been changed.

Comment

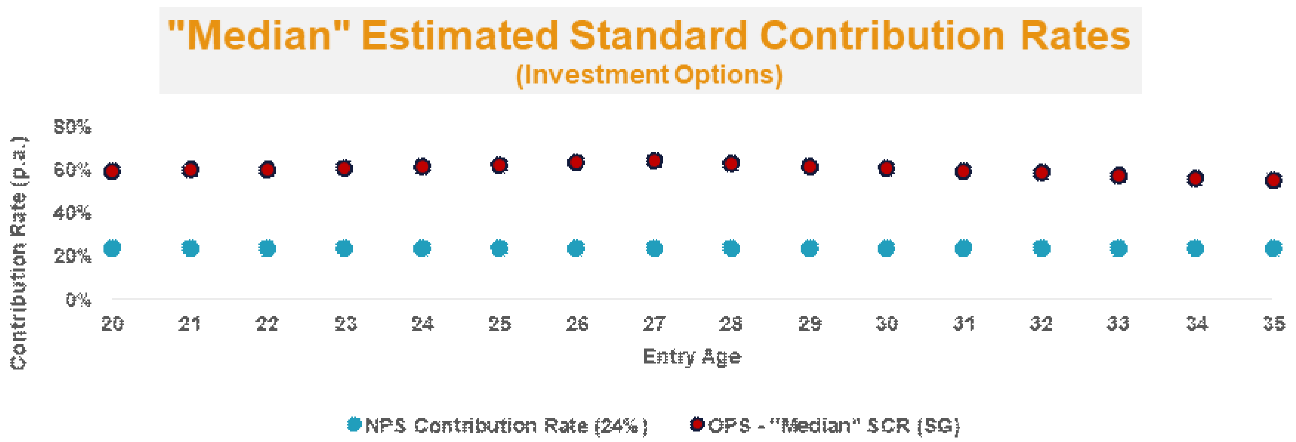
Assumptions used for long-term estimates are of course not a prediction. The actual circumstances will differ from the assumptions made. The results in our exercise will be most sensitive to assumptions (and more importantly the difference between) such as the discount rate and salary growth, discount rates and inflation/pension escalation. It is the difference between those assumptions that influence the sensitivity of the results more than the nominal value of each assumption in isolation.

Findings – Part #1

If an employee joined OPS today, then what would be an estimated standard contribution rate, using an entry age methodology and how does that compare to the current contribution into NPS (we take 24% of basic salary plus DA).

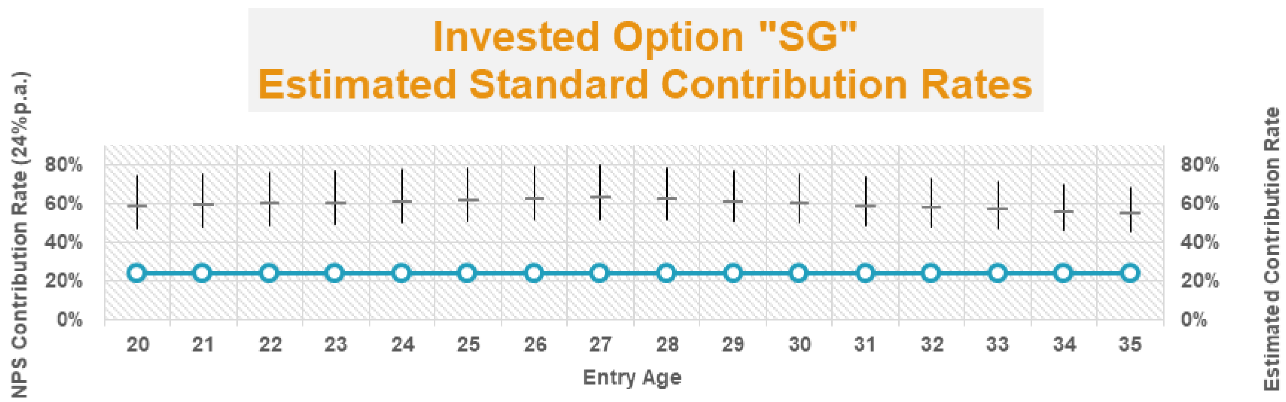
More specifically, with our assumptions, we demonstrate the confidence interval within which the NPS contribution rate fits into the entry age OPS estimated contribution rates.

The graph below shows our estimated standard contribution rates for the OPS. The first observation is just how large the numbers are. The median ranges from 55% of pensionable salary to 64%. The 24% being contributed into the NPS is far behind.



Graph 5: Median Estimated Standard Contribution Rates

We were hoping that the 24% would fall in some confidence interval but seems like an almost zero chance of that. The graph below shows the range of the estimated contribution rates. The vertical line represents 5% to 95%. The horizontal line is the median, which was also plotted in the above graph.



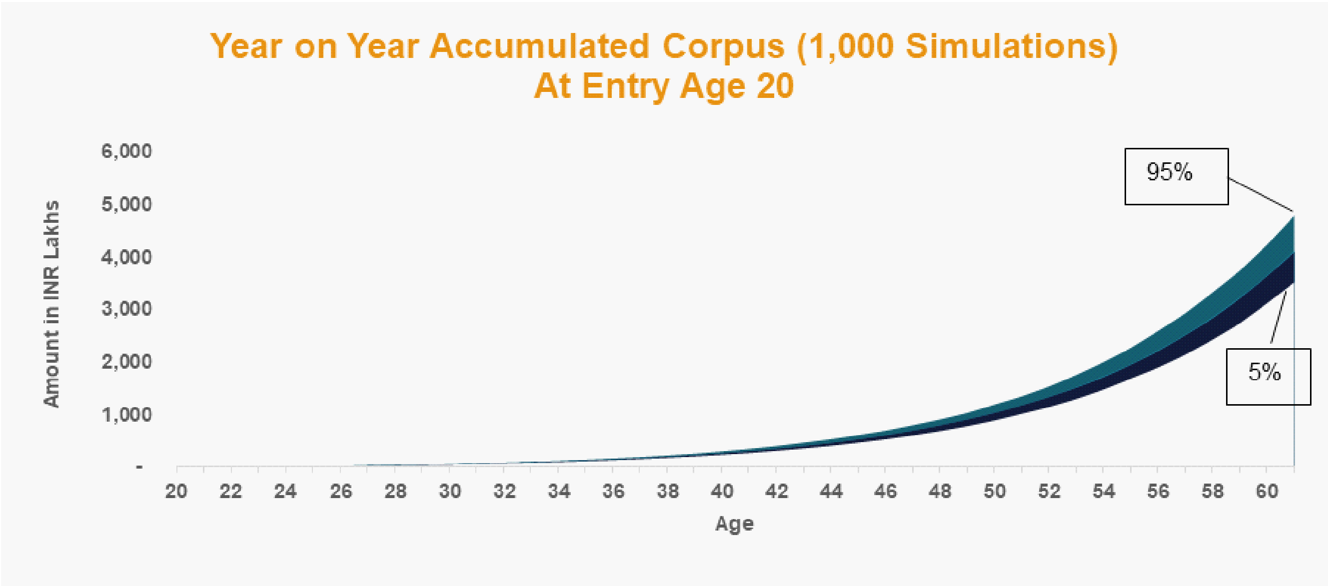
Note: The blue line represents the NPS 24% contribution rate

Graph 6: Range of Estimated Standard Contribution Rates

This sets the scene and basis to investigate what impact certain changes within the NPS could be utilised to increase the projected retirement income stream for individuals, then compared to the OPS.

To do that, we first projected sample lives' balances from entry age to retirement.

Focussing on the sample live age 20, below illustrates the projected NPS accumulated balance for such an individual. This assumes the investment performance in the same way as the discount rate for OPS used in Part I, i.e. the "SG" NAV performance. The difference in projected value is 36% between the 95th and 5th percentile by the time the individual completes age 60.



Graph 7: Accumulation of Corpus - Entry Age

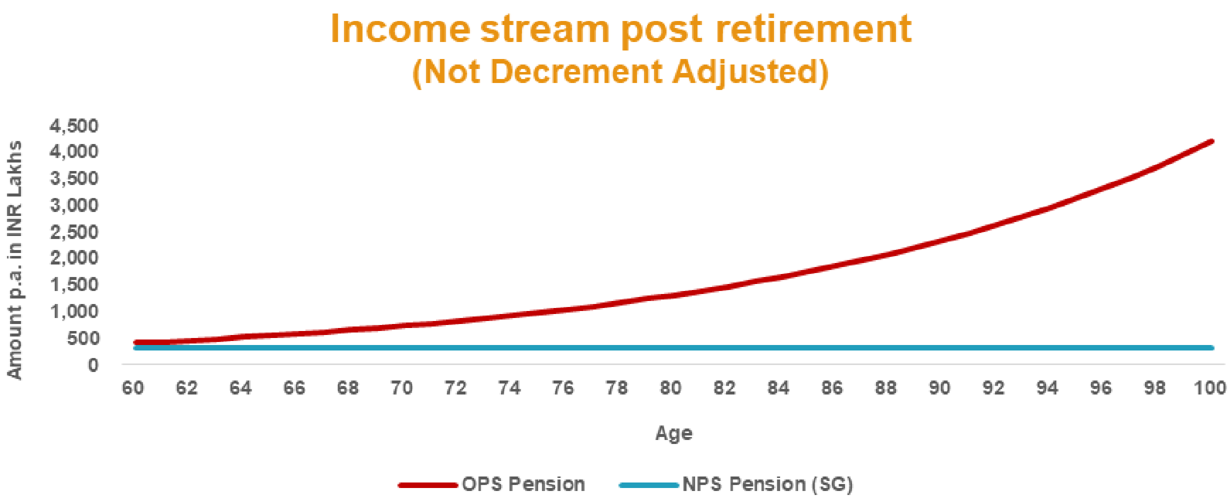
We then took the median value to move onto Part #2 for the income stream comparisons. In our sample life, the median value at age 60 is just over INR 400 million (INR 4,000 lakh). As mentioned earlier, the absolute values are less important given we are wanting to look at comparisons and we have an assumed salary for our calculations.

Findings – Part #2

Compare projected retirement income streams from age 60 under OPS and NPS along with the current default NPS annuity and fund management approach for the State employees.

How does that comparison change if we generated a range of potential NPS income streams by altering the pre-retirement asset allocation to the Lifecycle auto choice allocation?

Continuing from the previous section, we have used the median accumulated balance for our entry age 20 sample live and arrived at a projected income stream (post commutation). Having found out what we did in Part #1, the pattern and difference in income stream was to be expected. However, it is only when plotted like this does one see the significant role of the dearness relief in payment under OPS. The starting pension under NPS is about 20% lower than OPS. However, by age 80 that increases to about 75%.

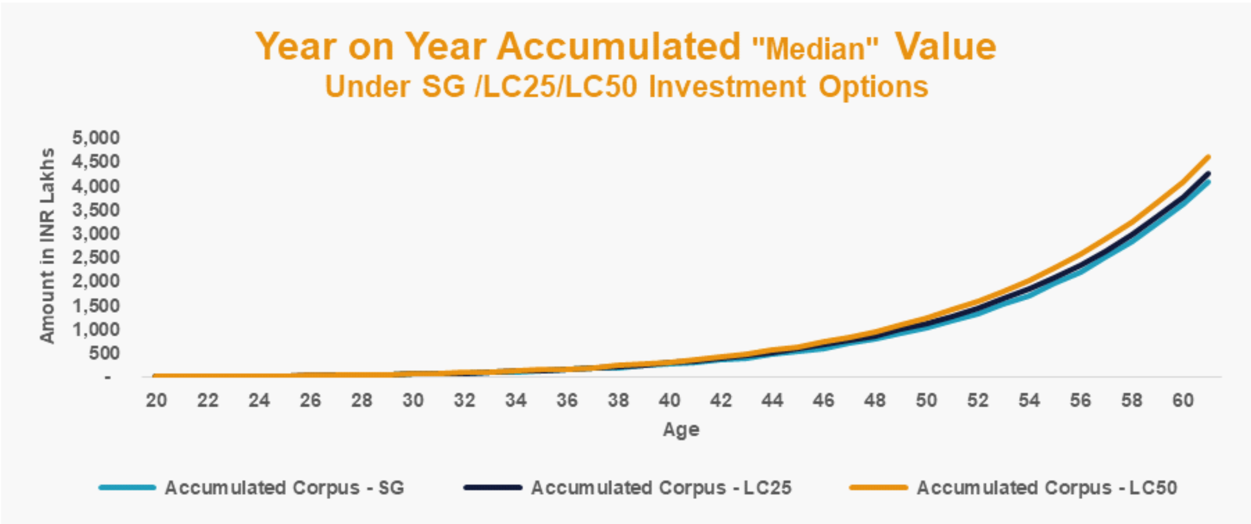


Note: The conversion of the NPS balance to pension is outlined in the methodology section and Appendix 3.

Graph 8: Post Retirement Income stream

But how about if we considered alternative asset allocations for the NPS funds?

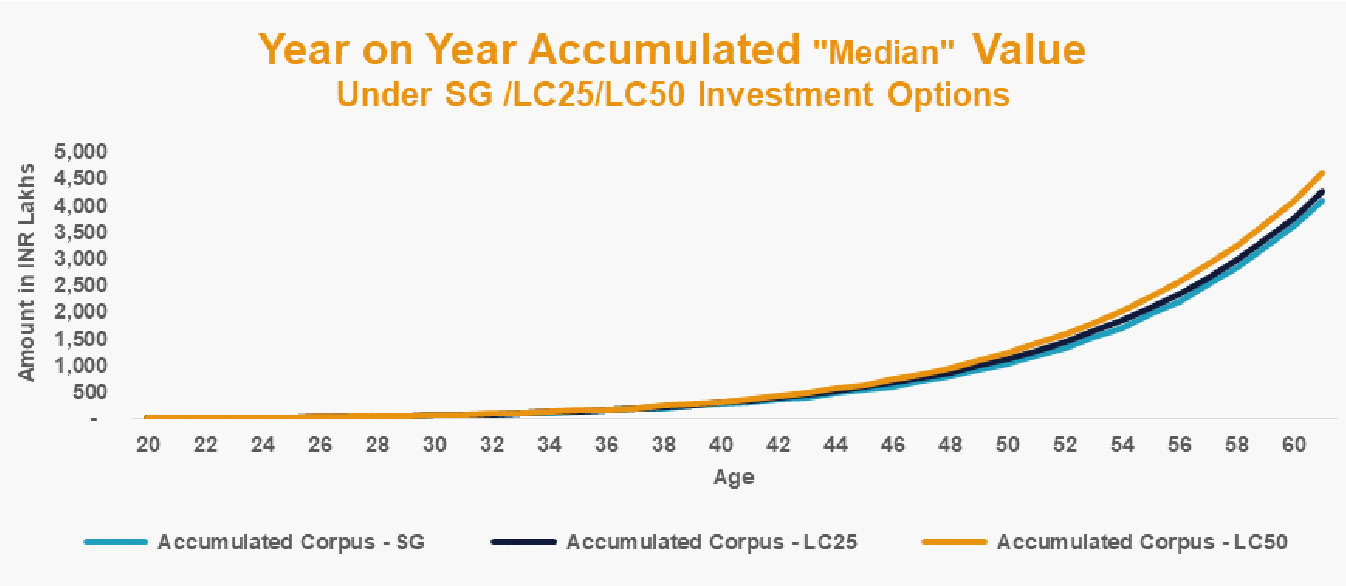
Below is a graph of the accumulated median balances under two alternative asset allocations, namely, Lifecycle 50 and Lifecycle 25, options that already exist for Central Government employees, and the all citizens NPS.



Graph 9: Median Accumulation of Corpus – Lifecycle asset allocation

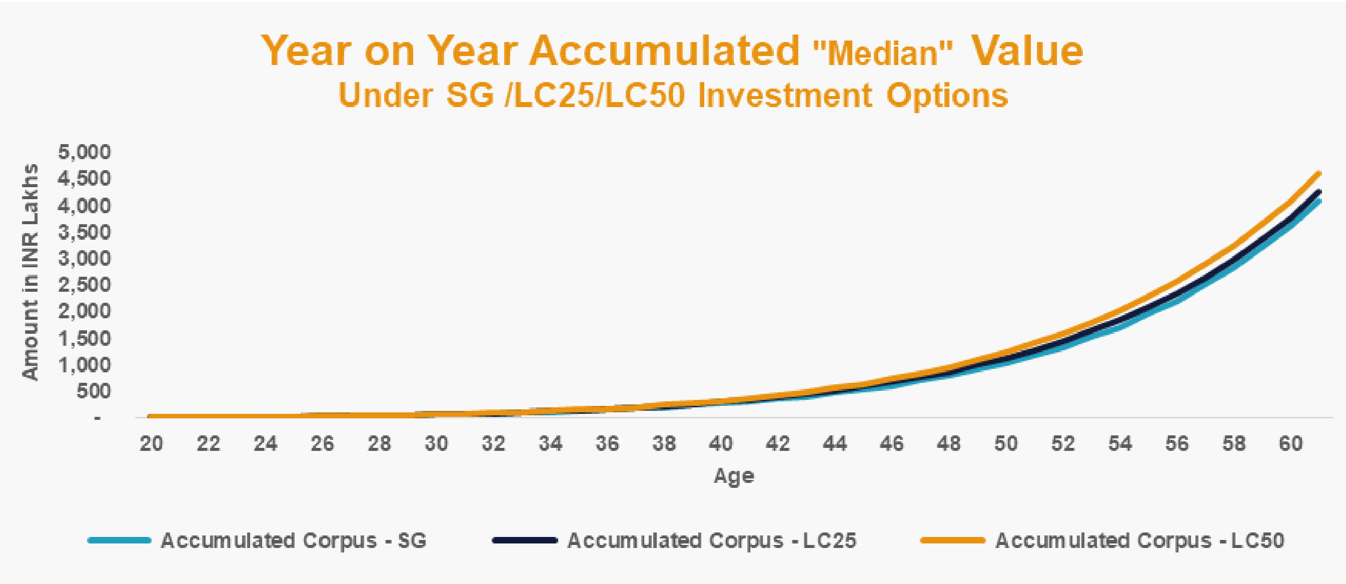
The values for LC25 and SG are very similar. This was largely expected given the similar underlying strong bias to Government Bonds and Fixed Interest in both allocations. We did expect greater difference under LC50. The difference at age 60 is about 13% when compared to the SG value. One reason the difference is not greater is probably because the LC50 has 50% in equities only until age 35, after which it reduces each year.

When we convert these into income stream, the additional pension from LC50 is the same 13%. In the overall context of the OPS comparison it means that asset allocation pre-retirement is not going to have a significant impact to bridge the gap to the OPS pension. We had expected a greater impact.



Graph 10: Post Retirement Income stream – Lifecycle asset allocation

We next introduced a mortality decrement adjusted graph of the income streams below. This so a fairer comparison can be made in Part #3 for the income drawdown streams.



Graph 11: Post Retirement Income stream – Decrement adjusted - Lifecycle asset allocation

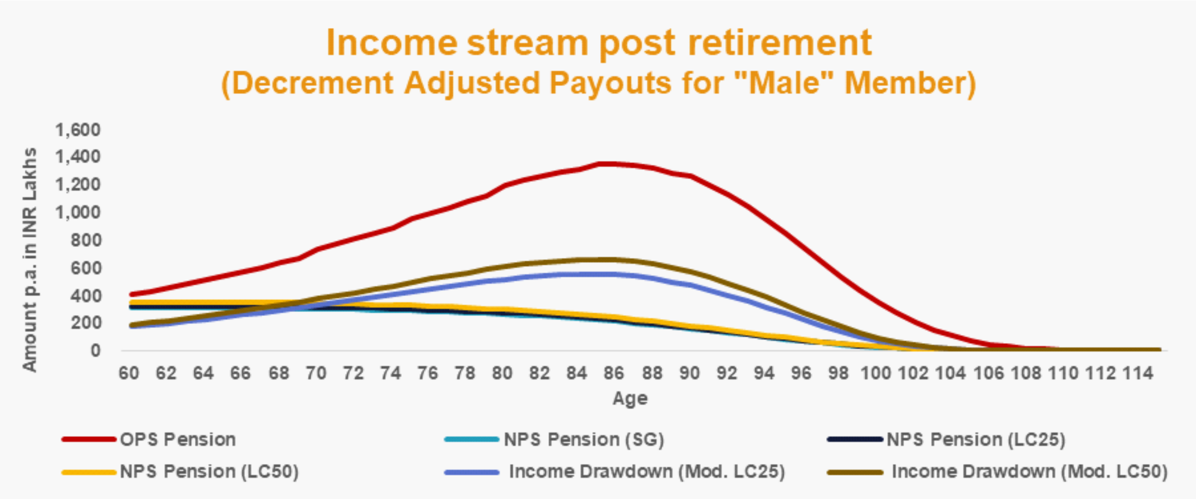
The shape of the OPS graph reflects the greater impact of pension increases up to about age 86. The mortality impact starts to take greater effect post that. The NPS graphs pretty much start reducing immediately as the core pension is flat.

FINDINGS PART #3

The findings so far show that we need look for something different to make any material impact on bridging the gap. The income drawdown approach ensures that a member's funds stayed invested after retirement. Our approach is outlined in the earlier section.

How does the comparison change if we generated a range of potential NPS income streams by altering the pre-retirement asset allocation was changed to the Lifestyle auto choice allocation, and the post-retirement income was from a lifetime income drawdown approach.

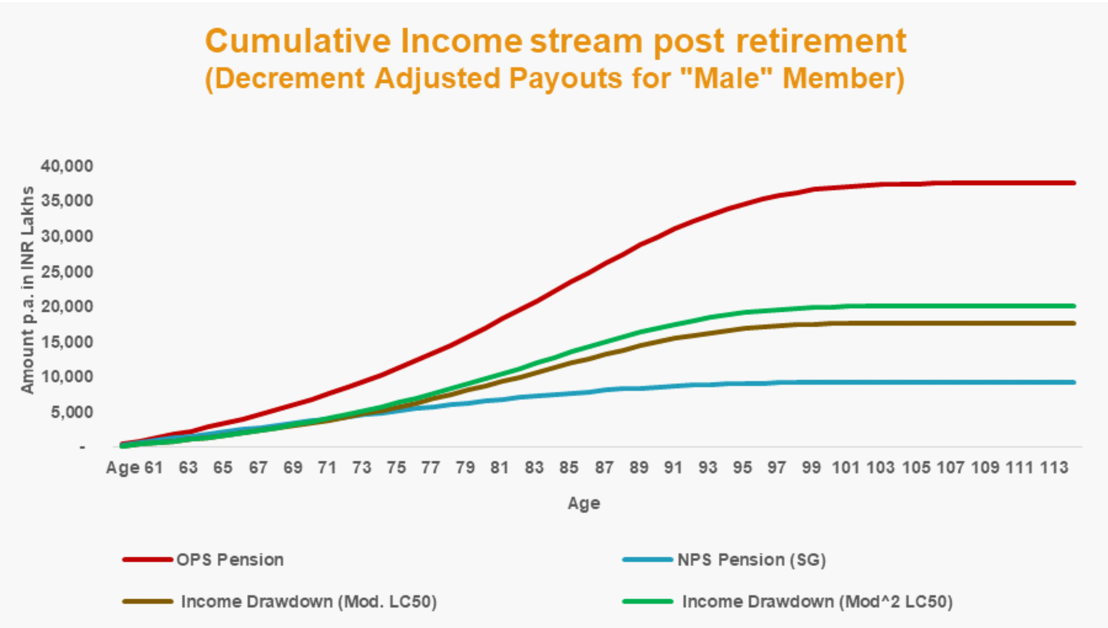
We have shown the income streams for different asset allocations. The LC25 and LC50 nomenclature has been updated to "Mod.LC25" and "Mod.LC50". This reflects the update we made to the lifecycle allocations, by deferring the change in pattern by 15 years. i.e. the lifecycle allocations start at age 50 instead of age 35.



Graph 12: Median Post Retirement Income stream – Decrement adjusted - Lifecycle asset allocation Modified

There is a material change in the income stream as compared to the original NPS pensions. Some observations using the Mod.LC50 scenario are as follows:

- The starting pension compared to the fixed NPS pension (SG) scenario is about 40% lower.
- However, it crosses at age 67 and then, at age 75, the cumulative amounts cross (see the next graph).



Graph 13: Median Cumulative Post Retirement Income stream – Decrement adjusted - Lifecycle asset allocation Modified

- By introducing the income drawdown in a modified lifecycle asset allocation, the cumulative income is now 50% of the OPS. It is 65% more than the traditional fixed annuity approach.

However, we still have a way to go, and it is seeming that, under the assumptions and approach for this paper, asset allocation will only go so far to bridging the gap to OPS. Income drawdown feature could however be a reasonable addition.

In a final roll of the dice, we have updated the LC50 and deferred for another 10 years. i.e. the lifecycle starts at age 60. Up to then one is invested 50%-E/30%-C/20%-G. Referred to as “Mod^2 LC50”. This adds another 13% on the cumulative income stream from the Mod.LC50.

In the decumulation phase, there is a further possible source of extracting potentially higher income from the same accumulated balance at retirement. The next section briefly introduces the concept of longevity risk pooling.

INTRODUCTION TO LONGEVITY RISK POOLING

Globally, defined contribution schemes are suffering from the fact that pensions generated from the balances are not at the level members’ expectations. There are a variety of reasons for this which we will not address directly here. However, designing decumulation payouts that have potentially better outcomes is a strong motivator for global pension experts to help in redress part of the situation. Whichever side of the debate one stands, one must recognise there is an increasing perception that annuities are not as efficient value for the decumulation phase as they once were.

One area of pension design theory that has seen a recent emergence, amongst experts, about longevity risk pooling designs in the decumulation phase. Generally called, Modern Tontines .[9]

A variation of the term is used in a paper by Price, Inglis, Ryder[10], namely “Value Annuities”.

The underlying concept is that a financial arrangement is created between members that form an asset pool. By being a member, one agrees to receive payouts while living but also forfeiting the account upon death. Forfeiture proceeds are then distributed among the surviving members. Payouts depend on investment performance and the mortality experience (mortality credits) of the membership pool.

When we began planning for this paper, one of our hypotheses was going to be to model a potential longevity pooling arrangement for a notional group of lives under the NPS and compare the anticipated income stream with OPS, and the income drawdown approach.

Based on our prior experience of the topic and having spoken to authors of the Value Annuities paper, we concluded that including the longevity pooling modelling for India may have limited application and value addition at the current time for the following reasons:

- The longevity experience portion is far less than the impact of the investment portion. We have covered the impact of investments/ drawdown in our study.
- Longevity pools in India would be very large. This means that the value addition on the income stream from the specific mortality experience within the pool is limited. In fact, the overall pool’s risk of population mortality changes could be more influential.
- The maximum potential value of these pools is when members forego all (or significant portion) of their balance on death. In India, this would be a large hurdle for members to buy into. Therefore, the core purpose of the longevity pooling concept diminishes greatly.
- Another core characteristic of such schemes is that the income stream can be variable. However, some minimum level of guaranteed

9 CFA Institute Research Foundation / Brief - Tontines - A Practitioner’s Guide to Mortality-Pooled Investments Richard K. Fullmer, CFA
10 Society of Actuaries, Variable Uninsured Life (Value) Annuities: Theory, Practice and Country Cases - Price, Inglis, Ryder (2021)

benefit can be incorporated into the design. It does though mean that there is again less funds "allocated / available" for sharing mortality credits. A recent article published in the United States[11] outlines an approach to overcome this challenge to some extent.

The modelling is also time consuming, especially if we were to consider an open-ended pool. We decided this further investigation would be a good theme for a future study within our profession in India. We believe it does warrant further studies, as there are also possibilities for insurance companies to create products utilising this concept too. Canada has recently allowed such pension fund savings products that incorporate longevity risk pooling features.[12]

We see a potential application for wider social security/pension schemes in India as well, such as Employees Provident Fund and Atal Pension Yojana.

One variation of the broader concept of risk sharing between the plan sponsor and members in a defined benefit scheme (which we have in the OPS) is to make them little like the Collective Defined Contribution Schemes which have started to operate in countries like Japan, UK in recent years. These essentially try to "target" a sustainable defined benefit formula but do not guarantee the defined benefit. Payouts are made from the fund and benefits are adjusted (or indexed increases adjusted) over time based on actual investment and mortality experience. Similar types of arrangements have existed for many years in Canada under unionised administered collective bargaining multi-employer industry wide plans.

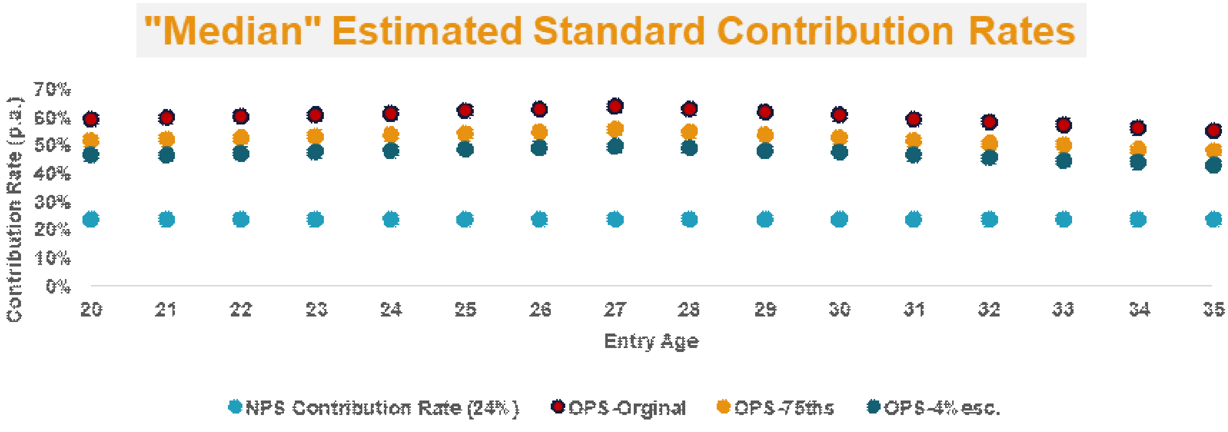
WHAT ELSE CAN BE DONE?

Our paper was not looking to alter the OPS benefits or contribution rates themselves. However, with the findings as they turned out to be we feel a short commentary on these is required. Mathematically, of course, the easiest thing to bridge the gap is to reduce the target OPS benefits or increase the NPS contribution.

As a final illustration we show the impact of two further scenarios.

- Reducing the OPS target accrual to 1/75th (rather than the existing 1/66th)
- Limit pension escalation of the target OPS to 4% p.a.

Below is the graph in Part #1 of estimated standard contribution rates with the new scenarios added. The average rate drops from 60% p.a. under the current benefits to 47% p.a. for the 4% p.a. pension escalation scenario. Of course, if we were to combine 75ths and 4% escalation then the rate would come down further to around 40-41%.

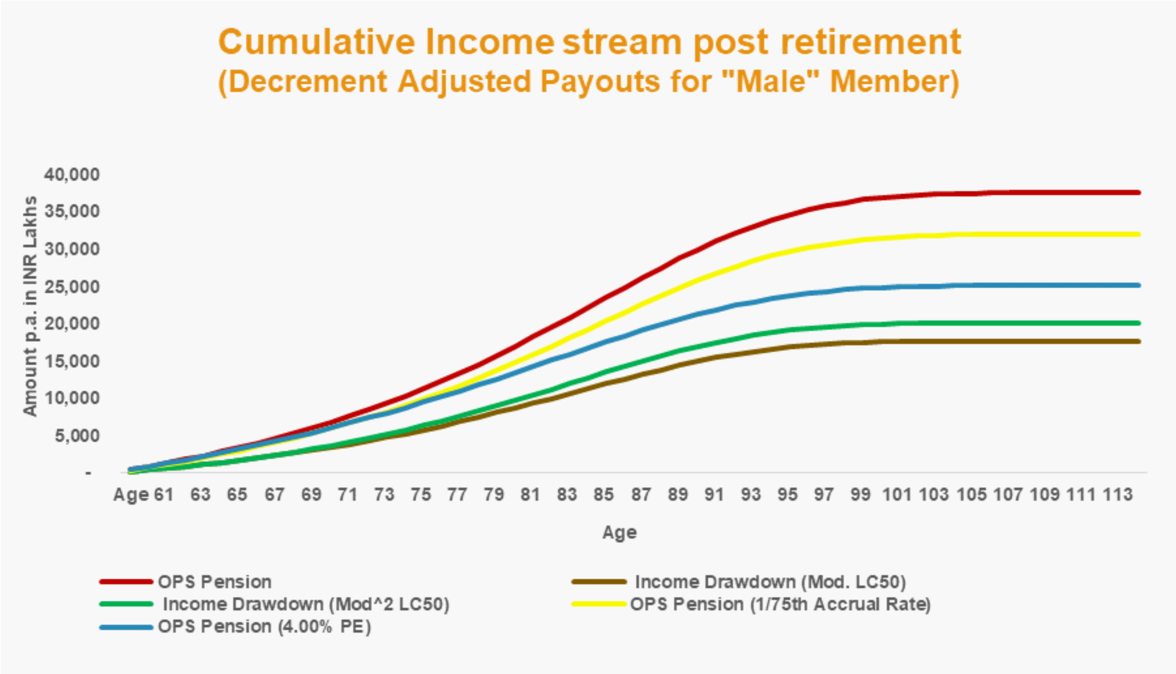


Graph 14: Scenarios-Median estimated Standard Contribution Rates

[11] Protected Modern Tontines: A New Approach for an Old Age Problem (2023) – David Blanchett and Gabriel Richards, published for Retirement Income Institute

[12] https://www.theglobeandmail.com/investing/adv/article-a-first-of-its-kind-fund-for-canadians-concerned-about-outliving-their/

To close, we illustrate the cumulative decrement adjusted income streams for our previous scenarios plus the new 75ths and 4% escalation scenario.



Graph 15: Median Cumulative Post Retirement Income stream – Decrement adjusted - Lifecycle asset allocation Modified

We think the comparison scenario in terms of feasibility to consider is the Modified LC50 (i.e. lifecycle start at age 50) with targeting OPS with a pension escalation limit of 4% p.a. The Mod. LC50 cumulative total is 68% of the OPS 4% escalation income stream.

Any remaining gap now could possibly be reduced by increasing the NPS contribution rate. A 30% (instead of 24%) will get us to within 20% of the OPS 4% p.a. limited pension escalation scenario.

Is that a gap that could be palatable by all sides?

The gap might be something that a government might be willing to fund in some way (i.e. top up or guarantee a minimum benefit). Andhra Pradesh announced a 50% of last drawn pensionable pay target minimum benefit under the NPS. This was rolled out from October 2023.[13] The government will top up employee pensions from the NPS to that extent, with also spouse's pension and dearness relief being secured in retirement through the top up. They have essentially converted their pay-as-you-go defined benefit scheme into a hybrid funded DC scheme.

[13] Andhra Pradesh Guaranteed Pension System Act 2023

Summarising commentary

We embarked on a hypothesis that a few adjustments to the NPS assets, and the payout form could go a long way to bridging the negative perception of the NPS vs OPS debate. What we have found, even with our simplistic modelling that overcoming the perception won't be easy. However, we do believe there are a few things that could at least improve the situation. The base estimated contribution rates are just so high compared to the NPS contribution rate.

The most notable improvement to NPS income streams would be through introducing a lifetime income drawdown scheme. That alone could improve the cumulative estimated income stream by 65% as compared to the current market annuity rates. This includes a modified lifecycle 50 asset allocation.

It was interesting to see that changing the asset allocation alone from the current SG asset allocation to, even with the current LC50, only saw an improvement in income stream of about 13%.

From an OPS comparative benefit, the major component affecting the cost is the post retirement escalation. Having a limit e.g. 4% p.a. can play a significant role in bridging the perception gap.

As we mentioned, our paper is designed to provide a starting point for further studies. The following is a list of examples where we feel our approach and models could be refined in further studies.

- Future asset returns could be modelled using detailed economic scenario generators calibrated for India
- One could incorporate projections of annuity rates offered by the providers for annuity conversion of NPS balances, inline with asset value movements with an assumed matching portfolio
- Future mortality improvements to be incorporated
- A comprehensive study of the potential impact of including longevity pooling into the income drawdown option.

ACKNOWLEDGEMENTS

1. Thank you to Himani Naik and Jill Morakhiya at K.A. Pandit who supported our research and some of the modelling.
2. Thank you to Will Price and Evan Inglis, Authors of Variable Uninsured Life (Value) Annuities: Theory, Practice and Country Cases, November 2021, published by the Society of Actuaries who both spent some time with us to answer some technical and practical questions related to the longevity pooling concept they had modelled.
3. Thank you to Sumit Ramani, FIAI, CEO of ProtectMeWell for providing some critical support related to the income drawdown modelling that saved us so much time and effort.

REFERENCES

1. Pension Regulatory and Development Authority Pension Bulletin October 2023 https://www.pfrda.org.in/myauth/admin/showimg.cshtml?ID=2795
2. Fiscal Cost of Reverting to the Old Pension Scheme by the Indian States – An Assessment by Rachit Solanki, Somnath Sharma, R. K. Sinha, Samir Ranjan Behera and Atri Mukherjee https://rbidocs.rbi.org.in/rdocs/Bulletin/PDFs/02AR18092023A646C561E9BC438AA67BFD934A58672E.PDF
3. United Nations, Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022, Online Edition.
4. Handbook of National Pension System Statistics 2023, PFRDA https://www.pfrda.org.in/myauth/admin/showimg.cshtml?ID=2805
5. Annuitants Table on Institute of Actuaries of India website: https://actuariesindia.org/sites/default/files/2022-05/Indian_Individual_Annuitants_Mortality_Table(2012_15).pdf
6. CFA INSTITUTE RESEARCH FOUNDATION / BRIEF

   Tontines - A Practitioner's Guide to Mortality-Pooled Investments by Richard K. Fullmer, CFA
7. Society of Actuaries, Variable Uninsured Life (Value) Annuities: Theory, Practice and Country Cases - Price, Inglis, Ryder (2021)
8. Protected Modern Tontines: A New Approach for an Old Age Problem (2023) – David

Blanchett and Gabriel Richards, published for Retirement Income Institute

9. Andhra Pradesh Guaranteed Pension System Act, 2023 https://www.apteachers.in/2023/10/ap-gps-gazette-notification-act-no-31.html

TABLES

GRAPHS

APPENDIX 1

Transition Dates of State Government from OPS to NPS

2003	Himachal Pradesh
2004	Punjab, Rajasthan, Andhra Pradesh, Chhattisgarh, Jharkhand
2005	Madhya Pradesh, Manipur, Odisha, Assam, Gujarat, Uttar Pradesh, Goa, Bihar, Uttarakhand, Maharashtra
2006	Haryana, Karnataka, Sikkim
2008	Arunachal Pradesh
2010	Jammu and Kashmir, Nagaland, Meghalaya, Mizoram
2013	Kerala

Source: National Pension System Trust

APPENDIX 2

Summary of Key benefits schedule for OPS used in our calculations

The pension scheme under OPS benefits is as outlined below:

Benefit	Eligibility	Description	Formula
Service Pension	Qualifying service** > 10 years		Last basic pay*** x(Qualifying service (max. 33 yrs)/66) + DR as applicable
Enhanced Family Pension	In case of Death, where Service is not less than 7 yrs continuous service	For 7 yrs from death OR Attaining age of 65 yrs had he been alive, whichever is earlier	50 % of basic pay drawn + DR as applicable; But limited to Service Pension
Normal Family Pension	In case of death of Govt. servant. Service < 7 yrs	Starts on completion of Enhanced Family Pension period.	30 % of the last basic pay drawn + DR as applicable
Additional Quantum of Pension	Pensioner attains age of 70 yrs and above	Increases by 5% for every 5 yrs of age up to 100 yrs. Beyond 100 yrs it is 50 %	Age related applicable% (10% to 50%) of basic pension + DR thereon
Commutation*	In case of Retirement only	Optional on Retirement. (Almost all take it)	(Max 40% of Service Pension) x 12 x Multiplying/ Commutation Factor
Service Gratuity	If a qualifying service <10 years Either Service pension or Service Gratuity is possible, Not both.		As per Rule 45: Ranges from (1/2) to 8 *(1/8) of Basic pay for every 6 months of completed qualifying service.

Table 6: OPS Benefit Summary

*Restoration of commutation is considered to happen after 15 years.

**Qualifying service is taken as service rendered till valuation date with bonus service of 5 years. Qualifying service is then capped at 33 years

***basic pay means basic salary plus DA

Both the OPS and NPS have many features in their benefit design and are quite complex. We have summarised a comparative table of the key provisions that we have used in our calculations.

Points of Comparison	Old Pension Scheme (OPS)	National Pension System (NPS)
Pension benefit	Employee gets up to 50% of his last drawn basic salary plus DA at age 60	Pension amount determined as per accumulated corpus and annuity opted at age 60
Contribution	No pre funding contribution from employee and employer	10% (Basic + DA) by employee and 14% government
Pension increases	Dearness relief applied	None
Commutation facility	Up to 40% of eligible pension. Commuted pension restored.	No commutation. However, Employee can take up to 60% of pension wealth as lumpsum payment at the time of retirement
Death Benefit	Applicable per the formula, mostly 50% of employee's pension	In service – Return of fund to nominee Post Retirement – Depends upon the selection of Annuity plan

* Subtle differences may exist from State to State but the base we have used is from an actual State, as it was in 2020.

Table 7: OPS vs NPS Comparison

APPENDIX 3

Annuity Rates source and derivation for NPS

Quotations generated on 21 December 2023

Source: https://cra-nsdl.com/CRAOnline/aspQuote.html

- Age: 60
- Quotations provide annual pension amounts (shown below)
- Quote generated as Male and also for Female (shown below)
- 50% Male/Female rates used
- Spouse's assumed 5 years difference (female spouse younger than male employee and vice versa)
- Quotation generated for a dummy INR100,000 corpus
- Averages only include annuity services providers where quotation was supplied
- Derived value is the average of the above formula across all annuity service providers. It came to INR13.1 for each INR 1p.a. pension (average of column h below)

	Without Return of Purchase Price		Without Return of Purchase Price		Annual Annuity rate		Used comparison
	Male		Female		Average Male/ Female		
	Annuity for Life	100% Joint Life	Annuity for Life	Joint Life Annuity	For Single Life	100% Joint Life	JS+60%
	(a)	(b)	(c)	(d)	(f) = Average a, c	(g) = Average b, d	(h) = (g-f)*0.6 + f
Bajaj Allianz Life Insurance Co. Ltd	7,898.73	6656.37	7,898.73	7,023.80	12.66	14.62	13.84
HDFC Life Insurance Co. Ltd	8,830	7,610	8,480	8,000	11.55	12.81	12.31
ICICI Prudential Life Insurance Co. Ltd	7,476	6,436	N/A (excluded)	6,673	13.38	15.26	14.50
IndiaFirst Life Insurance Co. Ltd	7,891.00	6,245.00	7,891.00	6,355.00	12.67	15.87	14.59
Life Insurance Corporation of India	8,685	7,308	8,685	7,889	11.51	13.16	12.50
MAX Life Insurance Co. Ltd	8,850	7,293	8,463	7,773	11.55	13.27	12.59
SBI Life Insurance Co. Ltd	8,221.00	7,352.00	8,221.00	7,680.00	12.16	13.30	12.85
Shriram Life Insurance Co. Ltd.	9,302.00	8,041.00	9,302.00	8,524.00	10.75	12.07	11.54

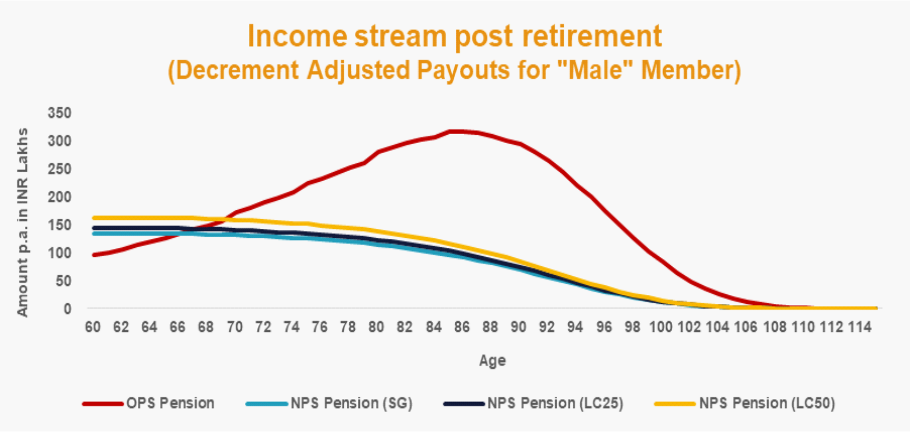
APPENDIX 4

Summary Findings using 8% p.a. salary growth.

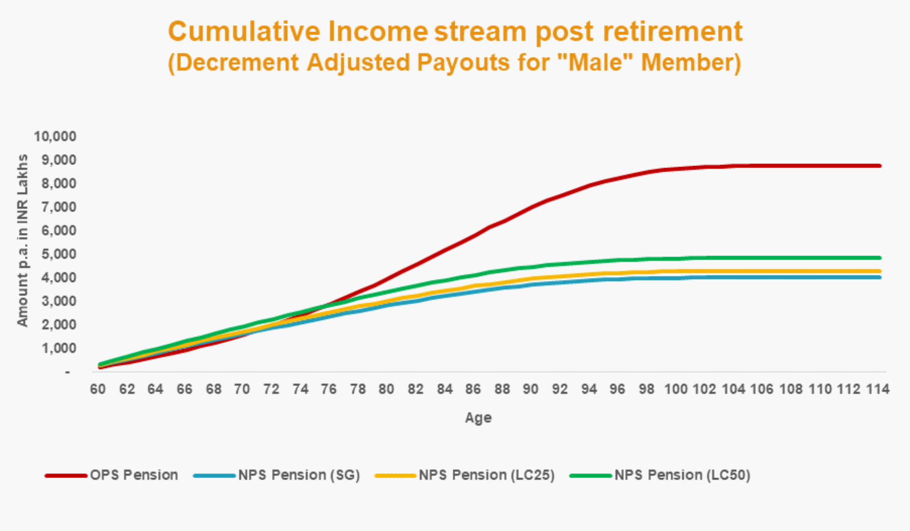
Estimated standard contribution rate calculated based on the Median Value with Salary Escalation Rate as 8.00% instead of 12%

The SCR now drops to a lot closer to the NPS, ranging around 32%-38%.

Taking just the 8% salary growth scenario for an age 20 entry age, the following comparison for the income stream (along the lines of Part 2 is below – see graph 11).



The cumulative graph looks as follows.



GCA 24

DATA, DISRUPTIONS AND THE ACTUARY

12, 13, 14 FEBRUARY, 2024 AT THE WESTIN POWAI LAKE, MUMBAI

PAPER IV

Navigating the insurance landscape: unravelling the impact of differences in profession on insurance choices

Navigating the insurance landscape: unravelling the impact of differences in profession on insurance choices

ASHISH ANTONY JAIN
SANDRA MARY K. S.

ABSTRACT

The decisions made by individuals regarding insurance are influenced by numerous factors, ranging from preferences associated with the insured's profession to economic considerations. The need to understand the complex interaction between profession, risk perception and decision-making to shape insurance preferences is obvious. The rationale of this study is that the insured's profession is more than simply a definition; it is a lens through which individuals assess, quantify, and navigate risk. From entrepreneurs who navigate the uncertainties of business ownership to workers who seek stability, each professional cohort brings unique perspectives and considerations to the insurance decision-making process. The study draws on a robust set of data encompassing demographic details, income levels and psychographic factors to construct a more in-depth understanding of insurance preferences. Preliminary findings indicate variations in risk tolerance in professional fields, which has an impact on product design and market segmentation. The research will contribute to a deeper understanding of the socio-economic foundations of insurance decision-making. In the outlook for the future, the implications of this research go beyond the scope of academic research. These resonate with the need for a more adaptable and consumer-oriented insurance sectors. By integrating the insights derived from this study, stakeholders have an opportunity to actively shape an industry based on the diversity of modern professional landscapes. We contribute not only to the theoretical foundations, but also to the practical development of rating factors based on an insured's profession.

KEYWORDS

Health insurance, Occupational perspectives, Demographics, Insurance coverage, Decision-making processes

CORRESPONDENCE DETAILS

*Correspondence to: Sandra Mary K. S., Master's candidate in Actuarial Science, University of Kerala, Student member, Institute of Actuaries of India (IAI), sandramaryks@gmail.com

1. INTRODUCTION

In the dynamic field of insurance, professionals face diverse choices that have a significant impact on the landscape. This paper examines the nuanced interaction between various industry perspectives, in order to reveal the profound impact these changes have on the choices made within the insurance sector. By examining how professional nuances form insurance landscapes, this paper attempts to illuminate the complexities that underpin decision-making processes. It emphasizes the importance of understanding and navigating these disparities, highlighting the need for a holistic understanding of how diverse expertise converges in the creation and improvement of insurance products. The choice of health insurance as the focus of our study among the various types of insurance stems from its unique intersection of personal well-being, financial security and social implications. Health insurance is a fundamental pillar, directly linked to the welfare of individuals and communities. Unlike other insurance that can address certain risks, health insurance encompasses a broader range of potential challenges ranging from routine medical care to unforeseen emergencies. This broad relevance and the complicated interaction of factors influencing health insurance decisions make it an ideal lens for unravelling the complicated dynamics of professional differences influencing insurance choices.

According to NITI Aayog annual report 2022-23; "Subsequent to the release of the NITI Aayog report 'Health Insurance for India's Missing Middle', the National Health Authority requested NITI Aayog to suggest a strategy to extend the health coverage to the 'missing middle' which is a broad category of population which lacks health insurance, positioned between the deprived poorer sections and the relatively well-off organized sector. Accordingly, a multi stakeholder committee, with Special Secretary, NITI Aayog as the Chair, was constituted and included members from Ministry of Health and Family Welfare (MoH&FW), National Health Authority, and other stakeholders including from the insurance sector.

The Committee is tasked with (i) devising a policy and strategy to extend or expand the health coverage to the missing middle under the ambit of Pradhan Mantri Jan Arogya Yojana (PMJAY) or independent of it; (ii) devising the criteria for identification of the missing middle as an individual or group and strategize a mechanism for their plausible enrolment or subscription for these; (iii) suggesting ways for effective distribution, raising consumer awareness of health insurance; and (iv) suggesting implementation pathways for extending coverage to missing middle. The Committee has organized several stakeholder deliberations, and the final report is under preparation."

The ongoing NITI Aayog study on India's missing middle, focusing on economically disadvantaged sections and the relatively affluent organized sector, further underscores the need for comprehensive health insurance strategies. The committee responsible for this study is actively working on devising policies, criteria, and implementation pathways to extend health coverage to the missing middle, with the final report currently in preparation. The insight derived from this ongoing study had contributed valuable perspectives to our research. Thus, our decision to focus on health insurance stems from the alarming reality that a significant portion of the Indian population lacks this crucial coverage. Despite its undeniable importance, many individuals in India are without this essential protection. This research aims to examine the intricacies of professional differences in health insurance preferences, emphasizing the urgent need for inclusivity in insurance access. By cantering our study on health insurance, we intend not only to address individual consequences but also to shed light on the pervasive gap in health coverage across the Indian demographic on a broader scale.

2. DATA COLLECTION AND METHODOLOGY

The dataset analysed was derived from primary data, meticulously collected through a well-designed online survey using google forms, which targeted 200 individuals spanning from 18 to 56+ from different districts of Kerala. In the survey sample, 50.5 percent of participants indicated they possess health insurance, with 37.3 percent relying on government-sponsored plans and 62.7 percent preferring private sector health coverage. The distribution of responses across various occupational sectors reveals a diverse economic landscape. The education sector takes the lead with 24.88%, underscoring its significant contribution to the overall survey. Business and financial sectors follow closely, each contributing 12.94% and 9.95%, respectively, reflecting their substantial roles in the surveyed population. Health services and social services sectors collectively contribute 20.90%, highlighting the importance of these fields. Meanwhile, manufacturing, hospitality and tourism, technological, and arts and entertainment sectors contribute moderately, showcasing a balanced representation of diverse professional backgrounds. The agricultural sector and pensioners, with 3.48% and 3.98%, respectively, emphasize the varied nature of respondents. This comprehensive breakdown provides insights into the multifaceted composition of the survey participants, capturing the broad spectrum of occupational backgrounds.

In the context of the bivariant analysis of our study, we further explored the associations within the data set. The investigation was extended to include an examination of chi square associations to assess the significance of relations between variables. This additional step improves the depth of our analysis, providing a nuanced understanding of the interaction between the different factors observed in the collected data. Employing the Statistical Package for the Social Sciences (SPSS) software, we conducted a thorough examination of the gathered responses. This analytical tool facilitated the extraction of valuable insights, revealing patterns and trends within the data. The strategic use of SPSS exemplifies our commitment to a precise and in-depth approach, utilizing technology to unravel the intricacies of each sector and age group. This integration of advanced statistical methods ensures a robust foundation for informed discussions within the respective fields.

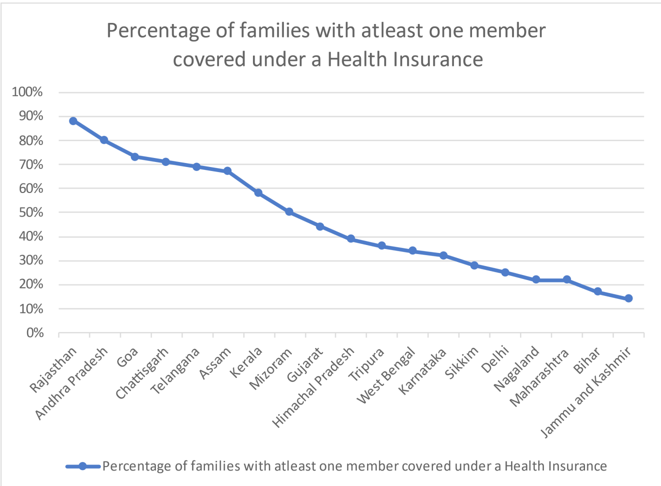
3.TRENDS AND INSIGHTS OF HEALTH INSURANCE IN INDIA

In a nation where health insurance coverage remains a critical concern, this paper looks into the complex network of insurance choices among professionals. By 2021, a staggering 514 million people, representing only 37 percent of India's population, were under the umbrella of health insurance schemes. Despite this, there is a significant gap, with about 400 million people fighting with zero access to such protection. Our exploration illuminates the dynamics of insurance preferences, with the aim of understanding the choices made by professionals in a landscape where more than 40 billion people navigate without health insurance security networks.

Gross Premium Collection of Health Insurance in INR Cr.	
FY15	20,096
FY16	24,498
FY17	30,392
FY18	37,029
FY19	45,532
FY20	50,752
FY21	58,237

(IRDAI Annual Report) Table 3.1

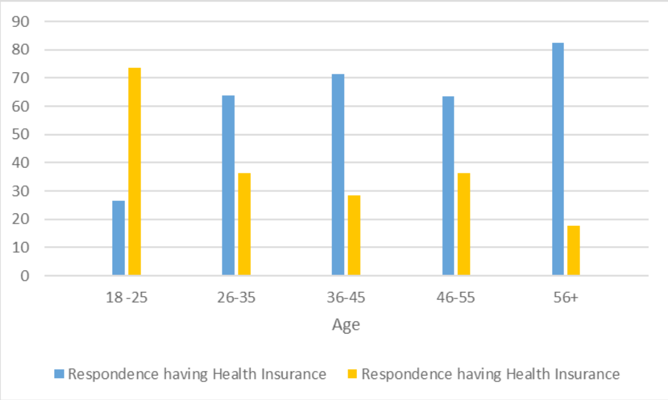
In exploring the health insurance landscape in different states of India, data from NFHS 2019-21 highlight significant variations. Rajasthan is leading with 88 percent of families with at least one member covered, while Jammu and Kashmir lags behind with 14 percent. This diverse spectrum emphasizes the impact of differences in an individual's profession on insurance choices.



(NFHS India Report 2019-21) Figure 3.1

4. HEALTH INSURANCE AND DEMOGRAPHICS

4.1 Association between age and health insurance uptake



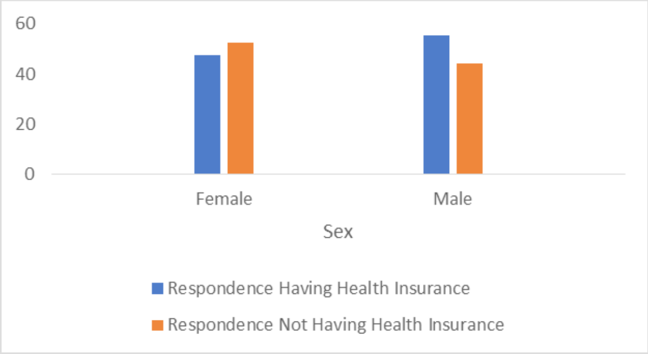
(Survey Results) Figure 4.1

The figure shows the association between age and health insurance prevalence, highlighting a progressive increase in coverage likelihood with age. In the 18-25 age group, a mere 26.5 percent hold health insurance, while a substantial 73.5 percent lack coverage, potentially influenced by factors like perceived lower healthcare needs, financial constraints, or limited awareness.

With advancing age, there is a consistent escalation in the probability of having health insurance. In the 26-35 age group, 63.8 percent possess coverage, and this percentage gradually rises in successive brackets: 71.4 percent (36-45), 63.6 percent (46-55), peaking at 82.4 percent for individuals aged 56 and above. This pattern suggests an increasing recognition of the significance of health coverage as individuals age, accompanied by a heightened vulnerability to health issues in older age groups.

The data underscores the necessity for targeted initiatives to augment awareness and accessibility to health insurance among younger individuals. Policymakers and insurers would benefit from deploying strategies tailored to address the unique concerns of diverse age groups, thereby fostering a more comprehensive and inclusive health coverage landscape across all demographics. This nuanced approach could contribute to bridging the coverage gap and ensuring a more equitable distribution of health insurance benefits across different age cohorts.

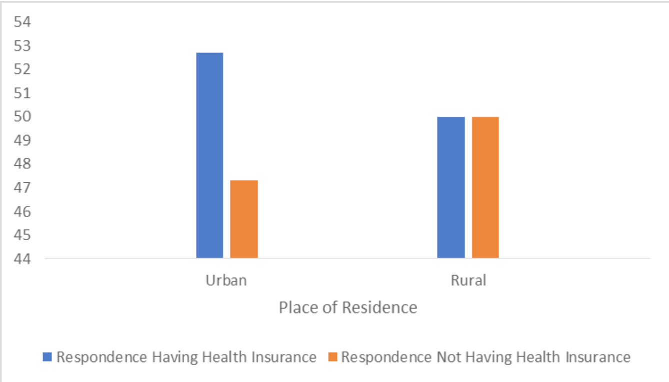
4.2. RELATION BETWEEN SEX AND HEALTH INSURANCE DECISION.



(Survey Results) Figure 4.2

The bar graph illustrates the percentage of people with and without health insurance based on gender. Looking at the data, we can see clear differences between males and females. Among females, 47.5 percent have health insurance, while 52.5 percent do not. This means that slightly more than half of females surveyed do not have health insurance. For males, the situation is different. A higher percentage, 55.6 percent, have health insurance, and 44.4 percent do not. These variations in health insurance coverage might be due to factors like the type of jobs people have, differences in income, or personal preferences in healthcare. It is important to dig deeper into these factors to create targeted plans and policies that can address the specific needs of both males and females. In summary, the data highlights the importance of considering gender-specific aspects when it comes to health insurance access. It suggests the need for customized strategies in healthcare policies to make sure that both males and females receive fair and adequate coverage.

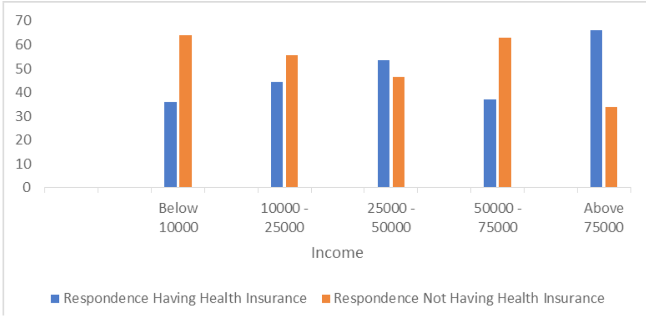
4.3. Interdependence of health insurance enrolment on place of residence



(Survey Results) Figure 4.3.

The bar graph visualizes the distribution of health insurance coverage among respondents based on their place of residence, distinguishing between urban and rural settings. The data reveals subtle distinctions in the health insurance status between these two contexts. In urban areas, 52.7 percent of individuals possess health insurance, while 47.3 percent lack coverage. This indicates a slight majority of urban residents having health insurance. Conversely, in rural areas, the graph shows an even split, with both 50 percent of respondents having health insurance and an equal percentage without coverage. The variations in health insurance coverage between urban and rural environments may be influenced by factors such as accessibility to healthcare facilities, economic conditions, and regional healthcare policies. Notably, urban areas exhibit a slightly higher prevalence of health insurance, possibly due to better access to healthcare services and increased employment opportunities. This graphical representation emphasizes the need for tailored healthcare policies that consider the unique dynamics of both urban and rural populations. Addressing factors contributing to differences in health insurance coverage is crucial for promoting equitable access to healthcare resources in both settings. Further exploration and a nuanced understanding of regional influences are essential for devising effective strategies to enhance overall health insurance coverage.

4.4 Linkage of income to the decision of health insurance adoption



(Survey Results) Figure 4.4

The figure illustrates the relationship between monthly income levels and the prevalence of health insurance, depicting the percentages of individuals with and without coverage across various income brackets. A clear pattern emerges, indicating a positive connection between higher income and an increased probability of having health insurance. Individuals earning less than Rs 10,000 display the lowest health insurance coverage at 35.9 percent, with 64.1 percent remaining uninsured. As income rises, the percentage of individuals with health insurance also increases. The Rs 10,000 – 25,000 income bracket shows an increase to 44.4 percent coverage, and this trend continues with 53.6 percent coverage in the 25000-50000 bracket. However, a notable deviation occurs in the Rs 50,000 – 75,000 income bracket, where coverage drops to 37 percent, suggesting a potential fluctuation or specific dynamics within this income range. The highest coverage is observed among those with incomes above Rs 75,000, with 66.2 percent having health insurance. This data underscores the influence of income on health insurance access, with higher-income individuals exhibiting a higher likelihood of being insured. Policymakers may need to explore targeted interventions to enhance coverage among lower-income groups, ensuring that health insurance becomes more accessible and affordable for a broader spectrum of the population. Additionally, understanding the peculiarities within the Rs 50,000 – 75,000 income range can inform specific strategies to address potential barriers in this particular segment.

5.THE CONNECTION BETWEEN JOB CATEGORY AND THE PRESENCE OF HEALTH INSURANCE COVERAGE

The results of the survey are summarized in Table 5.1.

Relation between Profession and Health Insurance (in percent)		
Profession	Respondents Having Health Insurance	Respondents Not Having Health Insurance
Education sector	52	48
Business sector	46.2	53.8
Financial sector	50	50
Health service sector	57.1	42.9
Manufacturing sector	50	50
Hospitality and Tourism sector	54.5	45.5
Social service sector	47.6	52.4
Technological sector	44.4	55.6
Arts and Entertainment sector	33.3	66.7
Agricultural sector	28.6	71.4
Pensioner	100	0

(Survey Results) Table 5.1

Getting medical help is important for everyone, but not everyone has the same access to health insurance. The above analysis delves into a detailed table outlining the distribution of health insurance coverage across various occupational sectors. The data reveals distinct patterns, shedding light on how different professions grapple with the divide between those with health insurance and those without. The table talks about many types of jobs, like teaching, business, finance, health services, making things, hospitality, social work, technology, arts, farming, and retired people. Each job group has a different number of people with health insurance compared to those without. This helps us understand the challenges different workers face.

In the education sector, the distribution of health insurance coverage is nearly balanced, with 52 percentage having coverage and 48 percentage lacking it. This sector may benefit from institutional policies or group health plans, making it more accessible for individuals to obtain health insurance.

The business and financial sectors present a contrasting pattern. In the business sector, 46.2 percentage have health insurance, leaving 53.8 percentage without coverage. This indicates a significant portion of individuals engaged in business activities lack health coverage, potentially due to the nature of entrepreneurship or the prevalence of small businesses struggling to provide comprehensive benefits. Conversely, in the financial sector, the divide is even at 50 percentage, which tends to have a balanced distribution of health insurance coverage, individual experiences may still vary based on specific companies, job roles, and local regulations. Additionally, economic factors, changes in healthcare policies, and other external influences can impact the relationship between the financial sector and health insurance coverage over time. In health service sector, expected to have high rates of health insurance coverage, exhibits a different trend. While 57.1 percentage have health insurance, a notable 42.9 percentage do not. Given the nature of the industry, health service workers may have greater awareness of the importance of health insurance, resulting in a higher coverage percentage. Similar to the business sector, the manufacturing sector's diverse landscape may contribute to an even distribution in health insurance coverage. The hospitality sector may experience fluctuations in coverage due to the prevalence of part-time or seasonal employment, affecting access to health insurance. In the social service sector, 47.6 percentage have health insurance, while 52.4 percentage do not, indicating a marginal disadvantage for those in social services. This sector might experience challenges in providing comprehensive health benefits, resulting in a slightly lower coverage percentage. In contrast, the technological sector exhibits a more pronounced divide, with 44.4 percentage having health insurance and 55.6 percentage without. The prevalence of contract or gig workers in the tech industry may contribute to this disparity. The arts and entertainment sector stands out with a significant gap; only 33.3 percentage have health insurance, leaving 66.7 percentage without coverage. This stark contrast underscores the vulnerability of individuals in creative professions regarding health insurance access. In the agricultural sector, a substantial 71.4 percentage lack health insurance, emphasizing a critical gap in coverage for those working in agriculture. The agriculture sector faces low health insurance coverage due to seasonal and temporary employment, financial constraints, lack of employer-sponsored plans, rural locations, and reliance on family labour, limiting accessibility. Conversely, pensioners exhibit a clear divide, with 100 percentage having health insurance, highlighting a system where retirees are provided with comprehensive health coverage.

In conclusion, the table paints a complex portrays of health insurance coverage across diverse occupational sectors. The variations in health insurance coverage among occupational sectors can be attributed to a combination of sector-specific employment practices, awareness levels, and the availability of employer-sponsored benefits. Addressing these factors through targeted policies and initiatives could contribute to a more equitable distribution of health insurance coverage across diverse occupational sectors.

6. OCCUPATIONAL PERSPECTIVES ON HEALTH INSURANCE CHOICES

6.1. Survey Findings on Reasons for Having Health Insurance

This analysis explores how different jobs relate to reasons for having health insurance. Various factors influence the decision to obtain health insurance, and the prospect's profession is one such determinant. In this analysis, we explore the correlation between occupation and reasons for having health insurance based on the provided data.

Relation between Occupation and Reasons for having Health Insurance (in percentage)					
Occupation	Respondence Not Having Health Insurance	For Tax Purposes	Financial Support for Health Emergencies	Due to a History of Hereditary Disorders	Suggestion From Others
Education sector	48	0	42	2	8
Business sector	53.8	7.7	26.9	3.8	7.7
Financial sector	50	0	40	0	10
Health service sector	42.9	9.5	47.6	0	0
Manufacturing sector	50	0	50	0	0
Hospitality and Tourism sector	45.5	9.1	36.4	9.1	0
Social service sector	52.4	0	47.6	0	0
Technological sector	55.6	0	33.3	0	11.1
Arts and Entertainment sector	66.7	8.3	25	0	0
Agricultural sector	71.4	0	28.6	0	0
Pensioner	0	0	100	0	0

(Survey Results) Table6.1

In the education sector, 48 percent of respondents do not have health insurance. Notably, none of them cite obtaining insurance for tax purposes. A substantial 42 percent acquire health coverage for financial support during health emergencies, reflecting a pragmatic approach to unforeseen medical expenses. A small percentage 2 percent considers a history of hereditary disorders, and 8 percent mention being influenced by suggestions from others. This sector emphasizes the importance of financial security during health crises, with minimal influence from external recommendations.

Moving to the business sector, 53.8 percent of individuals lack health insurance. A notable 7.7 percent obtain it for tax purposes, indicating a financial planning aspect for tax benefits. Additionally, 26.9 percent secure health coverage for financial support during health emergencies, aligning with the trend observed in the education sector. Interestingly, 7.7 percent mention being influenced by others, suggesting a more significant role of social factors in the decision-making process within this sector.

In the financial sector, 50 percent do not have health insurance. Similar to the education sector, none opt for health insurance for tax purposes. However, a substantial 40 percent acquire it for financial support during health emergencies. Interestingly, 10 percent cite suggestions from others, suggesting a moderate influence of peers in this sector. This indicates a recognition of the importance of financial security in times of health crises, with some influence from external recommendations.

The health service sector presents a relatively lower percentage (42.9 percent) of individuals without health insurance, possibly due to a better understanding of the importance of health coverage within the sector. A significant 47.6 percent secure health insurance for financial support during health emergencies, aligning with the overall trend. Notably, none mention a history of hereditary disorders as a motivating factor, indicating a focus on immediate health concerns rather than genetic predispositions.

The manufacturing sector mirrors the overall average, with 50 percent lacking health insurance. Similar to the financial sector, none acquire health insurance for tax purposes. However, a noteworthy 50 percent obtain it for financial support during health emergencies, emphasizing the sector's recognition of the need for economic stability during health crises.

In the hospital and tourism sector, 45.5 percent are without health insurance. A significant 36.4 percent opt for health coverage for financial support during health emergencies, and 9.1 percent consider a history of hereditary disorders. This suggests a dual focus on immediate financial concerns and long-term genetic health within the sector. Notably, no respondents mention suggestions from others, indicating a more independent decision-making process.

The social service sector reports a high percentage (52.4 percent) without health insurance. A considerable 47.6 percent secure health coverage for financial support during health emergencies, aligning with the trend observed in other sectors. Interestingly, none mention hereditary disorders or suggestions from others, indicating a specific focus on immediate financial concerns and a lower influence of genetic factors or peer suggestions.

The technological sector shows a higher percentage (55.6 percent) of individuals without health insurance. None opt for health coverage for tax purposes or due to a history of hereditary disorders. However, a significant 33.3 percent acquire it for financial support during health emergencies, emphasizing the importance of economic stability during health crises. Additionally, 11.1 percent cite suggestions from others, indicating a noteworthy influence of peers in this sector.

The arts and entertainment sector exhibits the highest percentage (66.7 percent) of individuals without health insurance. A considerable 25 percent acquire health coverage for financial support during health emergencies, aligning with the trend observed in other sectors. Notably, 8.3 percent mention suggestions from others, indicating a moderate influence of peer recommendations in the decision-making process.

In the agricultural sector, 71.4 percent of individuals lack health insurance. Similar to the education and financial sectors, none of them get it for tax reasons. A notable 28.6 percent secure health coverage for financial support during health emergencies, emphasizing the sector's recognition of the need for economic stability during health crises.

Pensioners, representing a unique category, universally acquire health insurance, with 100 percent stating financial support during health emergencies as the primary reason. This underscores the vulnerability of this group to health-related expenses and the importance of financial planning for unforeseen medical events.

The analysis of the relationship between occupation and reasons for having health insurance reveals diverse patterns across sectors. While financial support during health emergencies is a common motivator, the influence of tax benefits, hereditary factors, and peer suggestions varies. The data provides valuable insights for policymakers, insurance providers, and individuals to tailor health insurance policies and education efforts based on the specific needs and influences within different occupational sectors. Understanding these patterns can contribute to more targeted strategies for increasing health insurance coverage and promoting financial security across diverse professional domains.

6.2 Survey Findings on Reasons for Not Having Health Insurance

This analysis explores how different jobs relate to reasons for not having health insurance. In this in-depth look, we break down the survey results that show how people in different jobs think about and get health insurance. Whether it is because of money, trust, not knowing enough, or feeling like they would not get sick, everyone has got their reasons. Here we try understand why different jobs lead to different choices about health coverage.

Relation between Occupation and Reasons for not Having Health Insurance (in percent)						
Occupation	Respondence Having Health Insurance	Not Affordable	Not Worth the Money	Trust Issues with Insurance Companies	Lack of Aware-ness	You Believe That You Will Not Get Seriously Sick
Education sector	52	22	6	4	12	4
Business sector	46.2	7.7	3.8	23.1	15.4	3.8
Financial sector	50	25	5	5	10	5
Health service sector	57.1	4.8	4.8	14.3	14.3	4.8
Manufacturing sector	50	18.8	6.3	12.5	6.3	6.3
Hospitality and Tourism sector	54.5	36.4	9.1	0	0	0
Social service sector	47.6	14.3	4.8	14.3	19	0
Technological sector	44.4	22.2	0	0	33.3	0
Arts and Entertainment sector	33.3	16.7	8.3	0	41.7	0
Agricultural sector	28.6	28.6	14.3	0	14.3	14.3
Pensioner	100	0	0	0	0	0

(Survey Results) Table 6.2

First off, we notice that the numbers change a lot depending on the job. People in the education field seem to be more into health insurance, with 52 percent of them having it. On the other hand, those working in agriculture have a lower rate at 28.6 percent. This tells us that the type of job might have something to do with whether someone decides to get health insurance or not.

A good number of people in the business sector (7.7 percent) say they cannot afford health insurance. This makes sense because not all jobs come with big pay checks, and for some, health insurance can be expensive. It is a challenge that needs addressing if we want more people to have the coverage they need.

Trust issues are also a big deal. In the business and social service sectors, quite a few folks (23.1 percent and 14.3 percent, respectively) do not trust insurance companies. This lack of trust can be a hurdle. If people do not feel good about the companies providing health insurance, they might be less likely to sign up. So, building trust is key.

The survey points out that in some sectors, like arts and entertainment (41.7 percent) and tech (33.3 percent), many people feel like they do not know enough about health insurance. It seems like there is a need for more information and education in these areas. When people understand how health insurance works and why it is important, they might be more likely to get it.

In the health service sector, interestingly, some respondents (4.8 percent) do not think they will get seriously sick. This might be because they work in health-related fields and feel more in control of their well-being. On the flip side, retirees (pensioners) are all in, with 100 percent having health insurance. This shows that age and life stage play a big role in the decision to get coverage.

Each job comes with its own set of challenges when it comes to health insurance. For instance, the hospital and tourism sector had a higher enrolment rate (54.5 percent), probably because health and well-being are a big deal in these industries. Knowing the unique challenges each job brings helps us come up with plans that work for everyone.

For the people making the big decisions, like policymakers, this data is a goldmine. It shows where the problems are and helps us figure out how to fix them. If health insurance is too expensive for some jobs, we need to find ways to make it more affordable. If people do not trust insurance companies, we have got to work on building trust. And if there are jobs where folks just do not know enough about health insurance, education is the key.

In a nutshell, the relationship between jobs and health insurance is quite complicated. It is not just about money; it is about trust, understanding, and even feeling invincible sometimes. By looking at these survey results, we can see the bigger picture. It is time to come up with smart solutions that make sure everyone, no matter their job, can make the best choices for their health.

7. INFERENCES

A significant proportion in education sector cites financial support for health emergencies as the primary reason for having health insurance.
In business Sector diverse range of reasons, including financial support for health emergencies and tax purposes, are evident. The highest percentage claims health insurance for financial support during emergencies are from health sector. Unsurprisingly, 100 percent of pensioners have health insurance.

Across various sectors, a consistent need for financial support during health emergencies is apparent. This suggests a universal concern for safeguarding against unexpected medical expenses.

Sectors with lower percentages, such as the Arts and Entertainment and the Technological sector, could benefit from awareness campaigns highlighting the importance of health insurance. Understanding the diverse reasons within this sector suggests the need for customized insurance plans that cater to specific occupational requirements.

While moving to reasons for not having health insurance affordability emerges as a significant barrier, particularly in these sectors. Introducing more budget-friendly insurance options could address this issue. Trust issues with insurance companies is also notable. Transparency in policy terms and better communication can build trust. Lack of awareness is evident in Arts and Entertainment Sector and Technological sectors. Educational programs and easy-to-understand communication materials can bridge this gap. A significant percentage believes they would not get seriously sick. Educational efforts should emphasize the unpredictability of health events.

We recommend as below for increasing the health insurance coverage.

1. Tailored Plans:

Develop insurance plans tailored to the needs of specific occupational sectors. For instance, plans emphasizing tax benefits might be more appealing to individuals in the business sector.

2. Affordability Initiatives:

Introduce subsidized insurance plans or

government incentives for sectors facing affordability challenges, such as the Arts and Entertainment and Agricultural sectors.

3. Trust-Building Measures:

Insurance companies should focus on building trust, providing clear and transparent policies. Customer testimonials and reviews can be highlighted to instil confidence.

4. Awareness Campaigns:

Launch targeted awareness campaigns in sectors with low insurance coverage, utilizing various media to reach a diverse audience. Emphasize the importance of health coverage for financial security.

5. Educational Programs:

Implement educational programs in schools, colleges, and workplaces to increase awareness about health insurance and its benefits.

Increasing health insurance coverage requires a multifaceted approach. Tailoring plans to specific sectors, addressing affordability concerns, building trust, and implementing robust awareness and educational initiatives can collectively contribute to a higher percentage of individuals holding health insurance. By understanding the nuanced reasons behind decisions, policymakers and insurance providers can create more effective strategies to promote health coverage across diverse occupational sectors.

8. CONCLUSION

Our survey findings shed light on the prevailing landscape of health insurance in India, revealing that 50.5% of respondents possess health insurance, while 49.5% do not. Within this insured demographic, 62.7% prefer private insurance, and 37.3% rely on government plans. The primary reasons considered for not obtaining health insurance are identified as affordability issues, lack of awareness, trust issues etc. However, a consistent and significant factor is lack of awareness. While considering government initiatives such as Ayushman Bharat Pradhan Mantri Jan Arogya Yojana (PM-JAY), it stands as the world's largest health assurance scheme, providing a substantial health cover of Rs. 5 lakhs per family per year for secondary and tertiary care hospitalization. Targeting over 12 crores poor and vulnerable families, approximately 55 crores beneficiaries, PM-JAY strategically encompasses the bottom 40 percent of the Indian population based on the deprivation and occupational criteria of Socio-Economic Caste Census 2011 for rural and urban areas.

In summary, our survey underscores the pressing need for collaborative endeavours to enhance awareness and empower individuals in making informed choices regarding health insurance. Moving forward, emphasizing education and outreach will not only bridge the current information gap but also foster a more resilient society, where accessing healthcare is devoid of financial apprehensions. Through strategic communication and advocacy, we can aspire to a future where every eligible individual not only recognizes but actively utilizes the available health insurance options. This concerted effort aligns with the ambitious target of achieving insurance coverage for all by 2047, contributing to a healthcare landscape in India that is characterized by equity and inclusivity.

9. ACKNOWLEDGMENT

We extend heartfelt appreciation to our esteemed teachers for their unwavering support, guidance, and invaluable mentorship throughout our academic journey. To my peers, your collaborative spirit and shared knowledge have enriched the learning experience. Thank you all for being instrumental in shaping our educational paths.

10. REFERENCE

1. Insurance Regulatory and Development Authority of India (Annual Report)
2. National Family Health Survey, 2019-21
3. Maheshwari, R. (2023). Top Health Insurance Statistics of 2023, Forbes Advisor INDIA
4. Statista
5. PIB Government Press Release
6. PHD Chamber of Commerce and Industry
7. Ministry of Health and Family Welfare annual report 2022-23
8. NEETI AAYOG annual report 2022-23

Institute of Actuaries of India

INSTITUTE OF ACTUARIES OF INDIA

Unit no. F-206, 2nd Floor, "F" Wing in Tower 2,
Seawoods Grand Central Plot no R-1 Railway Station,
Sector 40, Seawoods, Navi Mumbai, Maharashtra 400706