

**LEVERAGING MACHINE LEARNING TO DECODE INSURANCE
PURCHASING DISPARITIES IN CANADIAN HOUSEHOLDS: A PCA APPROACH**

by

Chongrui Zhou

BSc., University of Northern British Columbia, 2009

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE (MSC)
IN
BUSINESS ADMINISTRATION

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

August 2024

© Chongrui Zhou, 2024

Approval Page

Abstract

This thesis investigates the factors influencing insurance spending among Canadian households, employing advanced machine learning techniques and Principal Component Analysis (PCA). This research develops an integrated predictive model to forecast household expenditures on life, health, and auto insurance, incorporating a comprehensive range of determinants such as household characteristics, economic conditions, and regional differences.

Utilizing a robust dataset from the Survey of Household Spending (SHS) for the years 2010 to 2017, with 2019 serving as the validation year, the study applies PCA to manage high-dimensional data effectively, thereby enhancing the predictive performance of the machine learning algorithms used. The results indicate that the model predicts insurance expenditures with notable accuracy; however, it slightly underestimates life insurance costs with an actual expenditure of \$1,381 compared to the predicted \$1,263 while providing highly accurate forecasts for health insurance. The predictions for car insurance expenditures exhibit larger variances.

The findings highlight the substantial benefits of integrating PCA and machine learning to advance predictive analytics in the insurance industry. The study offers critical insights for insurance providers, policymakers, and consumers, laying a data-driven groundwork for strategic decision-making and policy development. Recommendations for future research include refining the predictive models and investigating additional variables that may influence insurance spending. This thesis not only contributes to the academic discourse but also provides actionable strategies to enhance the accuracy and efficacy of forecasting models in the insurance sector.

Table of Contents

Approval Page	i
Abstract	ii
Table of Contents	iii
List of Tables	iv
List of Figures	v
Glossary	vi
Acknowledgment and Dedication	ix
Chapter One	1
Introduction	1
Background	3
Problem Statement	4
Objectives	5
Chapter Two	6
Literature Review	6
Methodology:	11
Hypothesis and Predictive Model:	13
Data Collection:	15
Chapter Three	21
Data Analysis Narrative/ Empirical Results	21
Chapter Four	34
Conclusion:	34
Future direction:	37
Bibliography:	38
Appendix 1: Figures	41
Appendix 2: Tables	47

List of Tables

Table 1. Number of Observations and Corresponding Weighted Population Estimates.	47
Table 2. Eigenvalue Distribution After PCA Analysis of Life Insurance Spending	48
Table 3. Eigenvalue Distribution After PCA Analysis of Private Health Insurance Spending	49
Table 4. Eigenvalue Distribution After PCA Analysis of Car Insurance Spending	50
Table 5. 2019 Canadian Households Insurance Spending	51
Table 6. Mean Life Insurance Spending for Canadian Households: Averages Across Ten Distinct Groups	52
Table 7. Mean Health Insurance Spending for Canadian Households: Averages cross Ten Distinct Groups	53
Table 8. Mean Car Insurance Spending for Canadian Households: Averages Across Ten Distinct Groups	54
Table 9. Canadian Household Data: Post-VariMax Rotation Analysis for Life Insurance Holders	55
Table 10. Canadian Household Spending on Private Health Insurance: A Post-VariMax Rotation Analysis	57
Table 11. Canadian Household Data: An Analysis Post-VariMax Rotation for Car Insurance Holders	59

List of Figures

Figure 1 Eigenvalue Distribution After PCA Analysis of Life Insurance Spending	41
Figure 2. Mean Predicted vs. Actual Life Insurance Spending Across 10 Groups in Canada, 2019	42
Figure 3. Eigenvalue Distribution After PCA Analysis of Private Health Insurance Spending	43
Figure 4. Mean Predicted vs. Actual Private Health Insurance Spending Across 10 Groups in Canada, 2019	44
Figure 5. Eigenvalue Distribution After PCA Analysis of Car Insurance Spending	45
Figure 6. Mean Predicted vs. Actual Car Insurance Spending Across 10 Groups in Canada, 2019	46

Glossary

Component: a "component" is a derived variable created by combining the original variables of a dataset in a way that maximizes variance captured, used especially in techniques like Principal Component Analysis to simplify data complexity.

Couple With Kid(S): A subcategory of household type focuses on describing the nature of households based on the relationships between members present at the time of the interview. This classification helps in understanding the familial and social dynamics within different living arrangements.

Couple With Other: Couples with other related or unrelated persons. A subcategory of household type focuses on describing the nature of households based on the relationships between members present at the time of the interview. This classification helps in understanding the familial and social dynamics within different living arrangements.

Female Cloth Spending: This category includes clothing, footwear, accessories, watches, and jewelry for women and girls aged four years and older.

Higher Level of Edu: Highest level of educational attainment of the reference person or the spouse.

Loading: The loadings, from a numerical perspective, are equivalent to the coefficients of the variables and offer insights into which variables contribute most significantly to the components.

Lone Parent: Lone parent household with no additional persons. A subcategory of household type focuses on describing the nature of households based on the relationships between members present at the time of the interview. This classification helps in understanding the familial and social dynamics within different living arrangements.

Man Cloth Spending: This category includes clothing, footwear, accessories, watches, and jewelry for men and boys aged four years and older.

Other: Occupied rent-free by the household is a subcategory of Type of Tenure. This classification refers to dwellings that are owned by the occupants who are currently paying off one or more mortgages. It indicates the type of tenure of the dwelling at the time of the interview.

Other HH Type: Other households with related or unrelated persons. A subcategory of household type focuses on describing the nature of households based on the relationships between members present at the time of the interview. This classification helps in understanding the familial and social dynamics within different living arrangements.

Own With Mortgage: Owned with a mortgage(s) by the household is a subcategory of Type of Tenure. This classification refers to dwellings that are owned by the occupants who are currently paying off one or more mortgages. It indicates the type of tenure of the dwelling at the time of the interview.

PCA Analysis: Principal Component Analysis (PCA) is a statistical method used to reduce the dimensionality of data sets, increasing interpretability while minimizing information loss. It achieves this by transforming a set of correlated variables into a smaller number of uncorrelated variables known as principal components.

Rec Car Insurance: insurance spent on Recreation vehicle

RDC: Research Data Center

SHS: Survey of Household Spending

VariMax Rotation: A statistical method used in and principal components analysis to streamline solutions and improve the interpretation of outcomes.

1 Person Household: A subcategory of household type focuses on describing the nature of households based on the relationships between members present at the time of the interview. This classification helps in understanding the familial and social dynamics within different living arrangements.

Acknowledgment and Dedication

First and foremost, I extend my deepest gratitude to my supervisor, Dr. Chengbo Fu. His guidance and support were invaluable when I struggled to find a suitable topic. Dr. Fu helped me arrange the project with RDC and Statistics Canada, providing the necessary resources and connections. His patience and encouragement were instrumental in calming my restless heart and steering me toward this thesis topic. His insightful feedback and continuous support throughout the research process have been crucial in shaping this work. His dedication to my academic growth has not only helped me complete this project but also inspired me to pursue further research with confidence.

I am profoundly grateful to my committee members, Dr. Kafui Monu and Dr. Fan Jiang. Dr. Monu's unwavering support throughout this journey, never giving up on me, has been a source of great strength. Dr. Jiang provided critical guidance and pointed me in the right direction when I felt lost. Their collective expertise and constructive feedback have significantly contributed to the completion of this thesis.

Special thanks to Dr. Chen Jing, who helped me start this journey. During this time, I have explored various fields, read numerous books, and kept updated with the latest news. I also experienced one of the significant financial crises firsthand. Although I couldn't systematically document this major event, being at the forefront of it greatly enriched my life. As a result, I began investing with real money, which ultimately funded my school tuition.

I also want to thank Dr. Chen Liang for his inspiration and encouragement, which led me to discover my path.

I extend my thanks to Statistics Canada and RDC for providing me with the opportunity to use the SHS database and for having staff on-site to support me and answer my questions.

I am deeply thankful to my family for their unwavering support throughout this process. Their encouragement has been my greatest strength.

Reflecting on this journey, I realize the profound changes within myself. This experience has transformed my expectations and understanding of education, shifting from merely teaching and learning to facilitating learning. It has broadened my perspective on life, moving from a black-and-white view to understanding that life is often a guessing game. While we may not know the absolute truth, there are ways to make better guesses and get closer to it.

Lastly, I want to thank the University of Northern British Columbia (UNBC) for providing me with all the support and opportunities to complete this project.

Chapter One

Introduction

The Canadian insurance landscape has distinct differences emerging in consumer purchasing behaviours compared to their American counterparts (Cacace & Schmid, 2008). This difference creates potential market opportunities and reveals previously hidden aspects of consumer decision-making processes. Given the important role that insurance plays in protecting the financial stability of households and the wider economy, it is essential to conduct a thorough analysis of these behavioural patterns (Nam & Hanna, 2019; Browne & Kim, 1993; Ye et al., 2022).

This study utilizes advanced machine learning techniques and Principal Component Analysis (PCA) to examine insurance spending among Canadian households. The objective is to create a predictive model that illuminates the factors that impact insurance purchasing decisions, providing a useful tool for forecasting future household insurance expenses. This study is valuable for a range of stakeholders, such as insurance companies, policymakers, and consumers in Canada.

Previous research in this domain has mainly focused on describing insurance spending patterns (Mapharing, M., Otuteye, E., & Radikoko, I. 2015; Browne & Kim, 1993; Bucciol & Miniaci, 2011; Campbell, 1980). However, these analyses have been limited due to the complex and multifaceted nature of the variables involved. Our research stands out by using machine learning and PCA to transform these intricate variables into a cohesive and predictive framework. This approach fosters a deeper understanding of the factors driving household insurance expenses, ultimately assisting in predicting insurance spending and enhancing decision-making within the industry.

The research holds significant implications for various stakeholders. For insurance providers, it offers valuable insights into consumer behaviour, enabling the customization of insurance products and the implementation of effective risk management strategies.

Policymakers and regulatory bodies can utilize the critical data generated by this study to make informed decisions, thereby fostering a more resilient and adaptable insurance marketplace.

Furthermore, Canadian consumers stand to benefit from a clearer understanding of the factors that influence their insurance choices, empowering them to make more informed financial decisions.

This study examines the insurance purchasing patterns of Canadian households across the life, health, and auto insurance sectors. By employing Principal Component Analysis (PCA) within a machine learning framework, we aim to accurately predict insurance expenditures and identify the underlying factors influencing these purchasing decisions.

The prediction of 2019 data produced the following key findings. The average life insurance expenditure of Canadian households was \$1,381, slightly exceeding the predicted figure of \$1,263, indicating a minor underestimation by the model. The forecast for the average expenditure on health insurance by Canadian households was highly accurate, with a prediction of \$1,608 closely aligning with the actual expenditure of \$1,594. The estimated car insurance expenditure by the model was \$2,048, while the actual spending was \$1,720.

These findings illuminate the factors influencing insurance choices and underscore the potential of machine learning to enhance predictive analytics within the insurance sector. The research offers valuable insights to stakeholders in the insurance industry by outlining spending patterns and highlighting discrepancies between projected and actual expenditures. This paves the way for more precise and data-driven decision-making processes.

It is crucial to emphasize the financial impact of our predictive model. By leveraging PCA, we estimate Canadian household insurance spending with a high degree of accuracy. Precisely forecasting future insurance expenditures translates to significant economic benefits for various stakeholders. The financial implications of our study extend beyond academic inquiry, delivering tangible benefits for both insurance companies and policymakers. Our model fosters a deeper understanding of the dynamics at play within the insurance market, consequently strengthening the sector's economic resilience. It promotes data-driven and efficient financial practices, ultimately contributing to the sustained health and vitality of the insurance industry.

Background

The Canadian insurance industry is a vital pillar of the nation's economy, offering essential risk mitigation tools for individuals and families. It's a diverse sector encompassing various insurance types like life, health, and auto insurance.

This industry thrives on high competition and a wide range of players. From large, multinational corporations with comprehensive insurance portfolios to smaller, niche companies specializing in specific products or customer segments, Canadian insurance caters to a varied clientele.

Life insurance safeguards policyholders' dependents financially in the event of the policyholder's death. Many insurers offer supplemental products like disability, critical illness, and long-term care insurance alongside basic life insurance policies. The life insurance sector is a blend of established national and international companies, alongside smaller players specializing in specific life insurance products or catering to particular demographics.

Private health insurance complements Canada's public healthcare system by covering services like prescription drugs, dental care, vision care, and allied health services that fall

outside the public system's scope. A diverse group of insurers, including life and health insurance companies, property and casualty insurers, and specialized health insurers, offer private health insurance in Canada.

Research like this study, which explores key trends and patterns in life, private health, and car insurance spending, plays a critical role in understanding and anticipating these upcoming changes. This type of research is particularly significant given the crucial role insurance plays in individual and household financial security, and by extension, the health of the Canadian economy.

Problem Statement

Cacace and Schmid (2008) highlight that in 2005, private insurance financing accounted for 36.6% of total healthcare spending in the USA, compared to only 12.9% in Canada. This trend has consistently shown lower private insurance spending in Canada from 1990 to 2005. Despite the close cultural and economic ties between the two nations, Canadian households exhibit differences in insurance ownership compared to those in the United States. This disparity suggests potential opportunities for growth in the Canadian insurance market. However, the reasons behind this gap remain unclear, particularly considering the various factors that may influence insurance decisions.

The role of housing characteristics is particularly noteworthy, as research suggests that factors such as suitability, income, education, and age can significantly influence spending (Mapharing et al., 2015). Browne and Kim (1993) analyze the relationships between various economic, social, and cultural factors and the demand for life insurance across countries, aiming to understand the determinants of life insurance consumption rather than predicting future spending. Bucciol and Miniaci (2011) analyze the distribution of risk tolerance among U.S.

households, finding significant heterogeneity influenced primarily by age and wealth, while other factors such as education, gender, and race show no significant impact. Their study focuses on understanding these relationships rather than predicting insurance spending. Campbell (1980) analyzes the relationships between economic uncertainties, particularly labour income uncertainty, and the demand for life insurance, deriving optimal insurance demand equations and emphasizing the influence of risk aversion and perceived insurance costs on household decision-making. This research predominantly focuses on comprehending these theoretical relationships rather than predicting insurance spending.

Our research addresses this gap by developing a machine learning-based model to predict insurance spending patterns among Canadian households. By incorporating a broad range of variables, this study seeks to comprehensively understand the factors shaping household insurance expenditure.

Objectives

This study aims to provide a comprehensive data-driven analysis of factors influencing insurance expenditure in Canadian households, focusing on life, health, and car insurance. These spending patterns are influenced by various factors such as household income, home value, urbanicity, and geographic location, creating a complex multidimensional space. To handle this complexity, the research leverages machine learning, particularly Principal Component Analysis (PCA), for dimensionality reduction and feature extraction. This enhances the robustness and predictive accuracy of the model. The research aims to decipher and model the intricate web of factors influencing insurance spending through this synergistic approach, providing a robust analytical framework that combines data-driven insights with predictive capabilities.

Understanding how households spend on insurance has significant implications for various stakeholders. For insurance companies, this research offers valuable insights for developing effective strategies and personalized insurance plans, tailoring coverage options to different income brackets, and informing marketing strategies to target specific customer segments. Policymakers and regulators can benefit by using the data to identify demographics with low insurance coverage, leading to policies that promote competition and innovation in the insurance sector, increasing access to insurance and ensuring a more equitable market. Consumers can make more informed decisions about their insurance needs, potentially leading to lower costs and improved coverage, thereby enhancing their financial security and peace of mind.

Financial advisors and investors can leverage these insights to offer more accurate guidance and inform investment decisions. Advisors can help clients anticipate future insurance costs and identify potential savings opportunities, while investors can use spending trends and future predictions for strategic investment choices. Finally, this research contributes to the broader academic field by providing a methodology and data foundation for further studies in insurance and household finance. It fosters innovation and deepens the understanding of these fields, potentially extending the findings to different contexts or countries.

The primary contribution of this research is to produce a predictive model practical for use by insurance companies and policymakers. This model aims to enhance strategic decision-making and policy development, ultimately benefiting the insurance market and its stakeholders.

Chapter Two

Literature Review

This study significantly advances our understanding of Canadian household insurance spending by transitioning from merely identifying influential factors to proactively predicting

future trends. This shift marks a monumental leap forward, enabling regulators, policymakers, and insurance companies to make data-driven, anticipatory decisions.

Prior research has predominantly focused on the American context. For instance, Bucciol and Miniaci (2011) analyzed the distribution of risk tolerance among U.S. households, finding significant heterogeneity influenced primarily by age and wealth. Similarly, Campbell (2006) explored the financial behaviours and decision-making processes of American households, examining empirical data and theoretical models to highlight the challenges and discrepancies between actual and ideal financial practices. Nam and Hanna (2019) found that higher risk aversion decreases the likelihood of single-parent households owning term life insurance but increases the likelihood of owning cash-value life insurance, with smokers less likely to own term life insurance but more likely to own cash-value life insurance. This study utilized data from the Survey of Consumer Finances (SCF) to analyze American households.

However, Canada's unique characteristics, such as its universal healthcare system and distinct approach to genetic information in life insurance (Knoppers & Joly, 2004), necessitate a dedicated examination. While historical Canadian studies (De Bromhead & Borowiecki, 2016) offer valuable insights, they may not accurately reflect the current landscape due to significant demographic and technological shifts. Mapharing et al. (2015). considered a broader range of factors in a more recent study but lacked a model for predicting future trends.

Calvet, Campbell, and Sodini (2007) consider various factors such as financial sophistication (wealth, education, pension contributions, liabilities) and demographic characteristics (age, employment status, household size, entrepreneurship, immigration status). Key factors influencing life insurance spending include higher income levels, increased education, and urbanization, all of which positively impact demand. Inflation can negatively

affect demand unless mitigated by concurrent economic growth and reforms. Social and demographic changes, such as smaller family sizes and an aging population, further drive the need for life insurance as a source of financial security (Hwang & Gao, 2003). Guiso, Haliassos, and Jappelli (2003) make it clear that insurance spending is a significant component of household financial portfolios in countries like the UK and the Netherlands. This is driven by insurance technical reserves. This trend is influenced by demographic shifts towards an aging population, institutional developments such as pension reforms, and government policies promoting private retirement savings through tax incentives. Collectively, these factors increase the reliance on insurance products for long-term financial planning.

According to Browne and Kim (1993), the primary factors influencing life insurance spending include the dependency ratio, national income, government social security spending, inflation, the price of insurance, predominant religion, education level, and life expectancy. Bucciol and Miniaci (2011) found that risk tolerance in U.S. households decreases with age and increases with wealth, while education, gender, race, and household size do not significantly influence risk attitudes. Their results are robust across different portfolio definitions and sample compositions.

Liu, Zhang, Chen, & Yang (2021) discovered that superstition, specifically the belief in the zodiac year, significantly influences economic behaviour among Chinese rural households. This belief leads to an 18.5% increase in life insurance spending during the household head's zodiac year, underscoring the role of cultural beliefs in shaping financial decisions. Utilizing data from the Chinese Household Finance Survey (CHFS) and the Peking University Digital Financial Inclusion Index (PKU-DFIIC), Ye, Pu, and Xiong (2022) found that digital finance significantly promotes household participation in risky financial markets. Digital finance achieves this by

reducing investment barriers, enhancing access to financial information, and increasing risk appetite. Furthermore, it reduces wealth and cognitive thresholds, thereby reflecting the inclusive nature of digital finance.

While previous studies have explored insurance spending, they lack comprehensive predictive models that consider a diverse array of independent variables. Further complicating the picture, regional variations within Canada, such as urban size and provincial regulations, may also influence households' insurance decisions (De Bromhead & Borowiecki, 2016). However, the contribution of these regional differences to predicting spending remains understudied.

By creating a predictive model that utilizes data from the Survey of Household Spending (SHS) and incorporates factors like internet access and tenure length, this study aims to substantially advance the understanding of insurance spending among Canadian households. Analyzing these variables will enhance our comprehension of the determinants of insurance spending and provide valuable insights into future trends. This will ultimately facilitate better decision-making by all stakeholders, including households, insurers, and policymakers.

To provide a well-rounded perspective, this section explores previous research that sheds light on factors influencing insurance spending and related financial decisions, even if they are not directly relevant to our specific methodology. Di Matteo and Emery (2002) investigated the relationship between wealth and life insurance demand in Canada. Campbell (2006) highlighted the critical need for improved data on household finances and pointed out common financial mistakes made by households. Law et al. (2013) examined the rise of private healthcare payments in Canada. Research by Browne and Kim (1993), Hwang and Gao (2003), Outreville (2013), Merkoulova and Veld (2022), Bucciol and Miniaci (2011), and others explored the connection between income, education, risk aversion, and insurance demand across various

countries. Mapharing et al. (2015) examined determinants of life insurance demand in Canada, while De Bromhead and Borowiecki (2016) analyzed the relationship between immigration and life insurance demand using historical census data.

Our model will assess the predictive power of a comprehensive range of factors categorized as demographic (province, household type, education level, urban size), housing-related (suitability, tenure length/type, house value, building age), financial (major income source, internet access), and lifestyle (spending on clothing/shoes, recreational car insurance). This paper introduces a novel Principal Component Analysis (PCA)-based approach to predict future insurance spending. Unlike previous Canadian insurance demand studies (Mapharing et al., 2015) that focused on influencing factors, our work leverages PCA to construct a robust forecasting model. This represents a significant advancement from earlier studies solely focused on identifying influential factors.

This study's innovative approach to predicting Canadian household insurance spending through the application of PCA builds on a well-documented foundation of PCA's efficacy across various domains. PCA's effectiveness in constructing predictive models across diverse disciplines, including healthcare and insurance, is well-established. Bro and Smilde (2003) provided clear quantitative and visual evidence of the impact of these preprocessing steps on the interpretation of principal components and the variance. For instance, Kanchan and Kishor (2016) effectively utilized PCA to reduce the dimensionality of their dataset, mitigate noise, prevent overfitting, and improve the interpretability and performance of machine learning algorithms for predicting heart disease and diabetes. This led to the creation of more efficient, robust, and accurate predictive models. Similarly, Wang and Wang (2015) demonstrated PCA's remarkable ability to enhance the accuracy of supervised machine learning algorithms in

predicting diseases. This approach holds particular significance for our research, providing a robust methodological framework for handling the complex, high-dimensional data inherent in analyzing Canadian household insurance expenditures. The successful application of PCA in a multitude of contexts underscores its adaptability and proficiency in deciphering multifaceted datasets, a crucial capability for understanding the intricate dynamics of insurance purchasing behaviour.

Methodology:

Research Philosophy

This study adheres to a positivist research philosophy, which emphasizes the acquisition of knowledge through objective, verifiable evidence. This aligns perfectly with our goal of developing a predictive model for Canadian household insurance spending. A positivist approach necessitates a rigorous and systematic data collection and analysis strategy.

Quantitative methods are central to this research design. We will utilize large datasets from reliable sources like Statistics Canada's Survey of Household Spending (SHS) database. This allows for high levels of accuracy and precision in our analysis. Statistical analysis of this quantitative data will enable the construction of robust predictive models.

Our focus on quantitative methods is a deliberate choice. While qualitative research offers valuable insights, a quantitative approach is better suited to our goals. It allows us to generate generalizable results and facilitates in-depth statistical analysis and mathematical modelling, which are crucial for building a strong predictive model.

Predictive modelling is a hallmark of positivist research, and it aligns perfectly with our methodology. This approach thrives on the large datasets and advanced statistical techniques that are integral to this study. In essence, our research philosophy, grounded in positivism,

emphasizes empirical, observable data. We leverage robust quantitative methods to achieve our objective of predicting Canadian household insurance spending with accuracy and generalizability.

Research Approaches and Strategies

The foundation of this research is a top-down deductive approach, which is fundamentally rooted in the tenets of existing theories. This approach begins with a general theory, formulating specific hypotheses. These hypotheses are then tested empirically, allowing the research to confirm, revise, reject, or refine the initial theory.

In this study, our deductive approach will start with the general theory of insurance spending, drawn from an extensive literature review and existing empirical studies. From this broad theory, we will develop specific hypotheses related to the predictors of Canadian household insurance spending. These hypotheses will then be tested using quantitative data, allowing us to examine outcomes and draw conclusions.

The use of quantitative data is integral to our research strategy. This method will enable us to test specific predictions and make generalizations based on the results. Quantitative data offers a degree of measurement precision and objectivity that aligns well with our deductive approach. The data will be collected and analyzed using robust statistical methods. This approach, often used in scientific research, enables us to draw conclusions and make inferences about the relationship between the variables under investigation. The statistical analysis will include techniques such as regression analysis, time series analysis, and machine learning algorithms, depending on the nature and structure of the data.

These statistical methods not only provide a means for testing our hypotheses but also allow us to build a predictive model for Canadian household insurance spending. This model will enable us to make projections about future spending patterns based on current and past data.

Similar to Kanchan & Kishor, (2016), our study divided the data set into two parts: Year 2010~2017 as a training dataset and year 2019 as a test dataset. The training dataset is used to feed the algorithms, allowing them to learn from this data. After the learning phase, the test dataset is used to evaluate the performance of the algorithms.

Our research approach and strategy rest on the principles of deductive reasoning and empirical observation. This framework, grounded in the traditions of scientific research, provides a rigorous and systematic means for testing hypotheses and generating reliable, generalizable findings.

Hypothesis and Predictive Model:

Our research posits that the deployment of Principal Component Analysis (PCA) on a 32-dimensional dataset, representing Canadian household insurance expenditure, will facilitate the extraction of a condensed subset of principal components. These components are anticipated to encapsulate a substantial portion of the dataset's variance. The employment of these principal components as input features is hypothesized to enhance the efficacy of our predictive model.

Our model utilizes PCA within a machine learning framework to predict household insurance spending. PCA is employed for dimensionality reduction, transforming the original high-dimensional data into a smaller set of uncorrelated components that capture the most significant variance. Data for this study is sourced from Statistics Canada's "Canadian Survey of Household Spending (SHS)" database spanning 2010 to 2017, with 2019 data used for validation. Variables considered include household income, education level, household size,

location, home value, length of tenure, and expenditures on life, health, and car insurance. The model hypothesis posits that PCA will extract principal components encapsulating substantial variance, thereby enhancing the predictive model's efficacy.

In practice, the model collects and preprocesses data, adjusting for inflation using the Consumer Price Index (CPI). PCA is then applied to reduce data dimensionality, focusing on the most relevant features. The dataset is split into a training set (2010-2017) and a test set (2019) to develop and validate the predictive model. Principal components derived from PCA are used as input features for machine learning algorithms to train the model, which is subsequently validated using the test set to assess predictive accuracy. The trained model predicts insurance spending for Canadian households, allowing for an analysis of key factors influencing insurance purchasing decisions.

Similar models and methodologies have been discussed in various literature. For instance, Bro and Smilde (2003) highlight the importance of centring and scaling in component analysis, while Kanchan and Kishor (2016) demonstrate PCA's application in disease prediction, showcasing its role in reducing data dimensionality and enhancing model accuracy. Studies like Buccioli and Miniaci (2011) examine household portfolios and implicit risk preferences, which are crucial for understanding insurance demand factors. Moreover, Mapharing et al., (2015) investigate determinants of life insurance demand in Canada, emphasizing comprehensive variable analysis. The integration of PCA with machine learning is further illustrated by Wang and Wang (2015), who explore its use in stock market prediction, underscoring PCA's effectiveness in improving predictive accuracy. These references underscore the broad applicability of PCA in various fields, including insurance spending analysis and predictive

modelling, demonstrating its robustness in handling complex datasets and making accurate predictions.

The selection of PCA is not arbitrary but is underpinned by several critical considerations. PCA is renowned for its proficiency in simplifying complex, high-dimensional datasets. This characteristic is particularly germane to our study, given the intricate nature of insurance purchasing behaviour, which is embedded in a multitude of life factors. Our choice of PCA is further justified by the following rationale:

Complexity of Insurance Purchasing Behavior: The decision-making process in insurance purchasing is complex, being influenced by an array of life factors. It necessitates an analytical tool that can sift through numerous variables to isolate those with the most significant impact on insurance purchasing decisions.

Enhancing Model Robustness and Generalizability: To enhance the robustness and generalizability of predictive models, it is crucial to focus on the most relevant components within the dataset. PCA effectively addresses the issue of numerous variables, ensuring that the model remains applicable and reliable in practical scenarios while mitigating overfitting.

The application of PCA will simplify the dataset and produce a robust and generalizable predictive model. This model will provide reliable insights into Canadian household insurance expenditure behaviours.

Data Collection:

This research will utilize data from the Canadian Survey of Household Spending (SHS) from 1997 to 2019. However, for this analysis, the dataset will be restricted to the years 2010 to 2017 and 2019. The rationale behind this choice is twofold: firstly, a significant redesign of the survey methodology in 2010 renders the data pre-2010 less comparable (Tremblay et al., 2010).

Secondly, data for the year 2018 is not available, necessitating a focus on the period from 2010 to 2017, with 2019 serving as a benchmark for comparison with our predictions.

Although the SHS data does not constitute a panel dataset, it closely aligns with Campbell's (2006) five ideal household data set criteria. It offers a representative sample of the population, provides highly accurate data, clearly distinguishes between asset classes, measures a comprehensive breakdown of wealth, and allows for tracking households over time.

For this study, the data will be adjusted to 2017 levels using the Consumer Price Index (CPI). This adjustment ensures that all the values are comparable in real terms, eliminating the effects of inflation over the years under consideration.

We aim to apply Principal Component Analysis (PCA) to generate a predictive model, using variables such as household income, highest level of education, size of the population center, location (provinces), household size, length of tenure, and the value of the house. The dependent variables will be life insurance, private health insurance, and car insurance spending. These variables were chosen for their potential impact on insurance spending, as indicated by previous research.

Due to the skewed distribution of insurance spending, with many Canadian households reporting no spending on diverse insurance types, the analysis will exclude non-insurance buyers. This approach aims to obtain a more accurate picture of insurance spending patterns among households actively participating in the insurance market.

Software

This study will employ STATA, a widely used statistical software package, for data analysis. STATA offers robust capabilities for data manipulation, statistical modelling, and visualization, making it suitable for planned analyses.

Microsoft Excel is used for organization tasks, such as formatting and preliminary data matching. However, all data will be ultimately converted to STATA format to ensure consistency and facilitate comprehensive statistical analysis.

Ethical Considerations

This research project strictly adheres to the Statistics Canada Research Data Centre (RDC) guidelines to ensure the privacy and rights of individuals. Our commitment to ethical data practices is reflected in the comprehensive framework provided by the guidelines, which guarantee that all research data used here has undergone rigorous ethical and legal review. Furthermore, transparency is paramount to our research approach. We will meticulously document our data collection, analysis, and reporting methodologies. Open access to these documents will allow for thorough scrutiny and validation of our findings, fostering trust and confidence in our research.

This study upholds the highest ethical standards. We are committed to protecting individual privacy and confidentiality, maintaining the integrity of the research process, and making meaningful and ethical contributions to the knowledge base on household insurance spending.

Data Preparation

To ensure the integrity and reliability of our study, meticulous data preparation was paramount. Our analysis utilized the Canadian Survey of Household Spending (SHS) data spanning the years 2010-2017 and 2019. For consistency in comparing data across these timelines, we adjusted all monetary figures to 2017 levels using the Consumer Price Index (CPI) for data analysis, and after making the prediction, we adjusted the prediction figure to several steps were undertaken to maintain data consistency:

Variable Consistency: Variable names were harmonized across datasets to ensure uniformity.

Data Merging: The pooled SHS data were integrated with the bootstrap data, creating a comprehensive dataset.

Data Transformation for STATA Compatibility included converting text entries to numeric values, ensuring proper formatting and labelling, and coding for missing values.

Appending Data: The pooled data from 2010 to 2013 was combined with data from 2014 to 2017. To ensure that this combined dataset accurately represented the entire Canadian demographic, we adjusted the weight variable by 50%. This adjustment was made to ensure that the data from 2010 to 2017 continued to accurately represent the entire population of Canadian households.

Intermediate Variable Creation: Variables such as pre-tax income, length of tenure, household size, urban size, insurance and clothing expenditure, house value, and recreational vehicle insurance underwent log transformations. Additionally, adjustments for CPI and data centring were essential for subsequent PCA analysis.

Refinements: Non-insurance buyers were excluded to prevent potential data skewing. Several categorical variables like education level, building age, and urban size were converted to continuous metrics for detailed analysis. Meanwhile, categories like provinces and major income sources were transformed into dummy variables.

After these refinements, we delved deep into the survey samples, examining variables like log income before tax, log length of tenure, education level, and urban size. This revealed insightful trends, such as rising life insurance purchases correlating with higher educational attainment.

Correlation and Variable Analysis

To construct a robust model, it was essential to understand the relationships between variables. We started with correlation tests, identifying and removing variables with high inter-correlations

to mitigate multicollinearity, which ensures a more reliable model. During the analysis of the dependent variable, we observed that a significant majority of the variables demonstrated strong relationships, indicating their importance for the model. We also transformed categorical variables into dummy variables to facilitate analysis and complied with RDC vetting requests by excluding certain variables. In managing outliers, especially in the income variable, we addressed data skew by capping income between \$5,000 and \$500,000, following RDC guidelines that recommended a thoughtful data cut-off and aiming to represent the majority of the household demographic accurately. Through meticulous data preparation, we have established a solid foundation for our predictive model, ensuring it is based on a clean, consistent, and insightful dataset.

Advantages of PCA analysis

Principal Component Analysis (PCA) has seamlessly integrated itself as a pivotal tool in multivariate data analysis, predominantly attributed to its adept capability to manage and decipher high-dimensional data. This is crucial in many real-world contexts, such as finance and social sciences, where data is ubiquitous and manifests across numerous dimensions or variables. PCA judiciously addresses these complexities by proffering a methodology that encapsulates this multivariate data in a concise, reduced form, judiciously preserving vital information. It ingeniously transmutes the original data dimensions into a fresh set of dimensions, termed 'principal components.' The orthogonality of these components is crucial, ensuring linear independence and offering a significant upper hand in sidestepping multicollinearity within predictive modelling.

In the contemporary era of data-driven decision-making, Machine Learning (ML) has burgeoned as an indispensable facet, especially in deciphering patterns within high-dimensional data, which

is often challenging to navigate. The integration of PCA and ML could potentially pave the way for more precise and computationally efficient predictive models by reducing the dimensionality of the data fed into ML algorithms, thereby mitigating the risk of overfitting and enhancing model interpretability.

The nuanced sophistication of PCA is manifested in its ability to prioritize these components astutely, anchoring this prioritization on the variance each component sequesters from the original data. A hierarchical structure emerges, wherein the first principal component sequesters the maximal variance, followed by the second, and so forth. This hierarchical variance capture enables a judicious reduction in the dimensions utilized for subsequent analyses or modelling, as typically, only the paramount components—those sequestering the maximal variance—are retained.

Leveraging the prowess of PCA allows researchers and analysts to distill their data structures, rendering them more navigable and interpretable and enhancing the efficacy and speed of subsequent modelling processes. Especially when integrated with ML algorithms, this amplifies the modelling process and ensures that models concocted upon this reduced data are less susceptible to overfitting, being stripped of redundant or insignificant dimensions. Thus, the resultant model is not only computationally efficient and robust but also adept at capturing the quintessence of the original data, tactfully navigating through the quagmires associated with high dimensionality.

In this paper, we explore the potent synergy between Principal Component Analysis (PCA) and Machine Learning (ML) in building robust and dependable predictive models for complex multivariate data, specifically focusing on Canadian household insurance expenditure behaviours. PCA elegantly transforms high-dimensional data into a set of uncorrelated "principal

components," capturing the most significant variance while mitigating multicollinearity. This dimensionality reduction unlocks a cascade of benefits: enhanced computational efficiency, reduced overfitting risk, and improved model interpretability. By channelling the power of ML through these streamlined components, we build models that not only achieve superior precision but also excel in robustness and generalizability. This advantage stems from PCA's ability to concentrate on key data features, leading to models that perform admirably in real-world scenarios. Our commitment to a quantitative analytical approach, fueled by credible data sources like the Survey of Household Spending, aligns perfectly with the data-driven philosophy of predictive modelling. This rigorous statistical and mathematical foundation ensures objectivity and generalizability, surpassing the limitations inherent in qualitative data and paving the way for insightful and dependable models that illuminate the intricacies of Canadian household insurance expenditure patterns.

Chapter Three

Data Analysis Narrative/ Empirical Results

The general predictive model of Canadian households' insurance spending is as follows:

$$PC_i = \sum_{j=1}^p \alpha_{ij} X_j$$

$$\hat{Y}_{PCA} = f_{PCA}(PC)$$

PC represents the principal components obtained from the original dataset.

f_{PCA} represents the predictive model function that uses the principal components.

\hat{Y}_{PCA} represents the predicted outcomes based on the predictive model.

α_{ij} are the loading for the original variables X_j

X_j are the original variables.

Life Insurance Predictive Model and Discussion

The life insurance predictive model is as follows:

$$\hat{Y}_{PCA,life} = \beta_0 + \beta_1 \cdot (PC_1) + \beta_2 \cdot (PC_2) + \beta_3 \cdot (PC_3) + \cdots + \beta_n \cdot (PC_n)$$

- β_0 is the intercept.
- β_i are the coefficients for the principal components PC_i

While exploring the life insurance portfolios of Canadian households, we encountered a statistical issue: high correlations (multicollinearity) between multiple independent variables. This phenomenon, where two or more predictors in a multiple regression model are correlated, poses a challenge as it can distort the interpretation and reliability of statistical results. To counteract this, we harnessed the power of Principal Component Analysis (PCA), a sophisticated and advanced statistical method often used to handle many variables, improve visualization, and reduce noise. This rigorous process allowed us to transform our primary set of 16 independent predictors into an expanded set of 33 distinct components. In doing so, we improved both the precision and robustness of our analysis, allowing for more nuanced insights.

At the outset, our methodological approach was anchored to Kaiser's (1960) widely acknowledged criterion. This involved adopting an eigenvalue threshold of 1 for component inclusion, a practice that has historically demonstrated efficacy in discerning principal components. However, as our investigation evolved and the intricacies of our dataset became evident, it became manifestly clear that solely relying on a limited subset of components inadequately captured the underlying complexities and inherent variance of our data. By recalibrating the eigenvalue benchmark to a more conservative 0.5, we succeeded in explicating a substantial portion, nearly 95%, of the variance. This methodological shift also had the advantageous effect of reducing the dimensionality by seven, streamlining our analytical processes. (Table 2, Figure 1)

(insert Table 2, Figure 1 here)

The dataset was not merely a collection of numbers; it encompassed 37,610 data points, which, when appropriately weighted, reflected a sizable population of 5,610,380. This number is not trivial; it is equivalent to as many Canadian households, providing a substantial sample size for our analyses (Table 1).

(insert Table 1 here)

Given the provided data, the Principal Component Analysis (PCA) was conducted to understand the underlying structures of various variables related to insurance. Specifically, the variable "couple with kid(s)" showed a strong loading of 0.6926 on Component 1, suggesting that it can be a pivotal variable in explaining the variance in this component. "Size of household" also presented a substantial loading of 0.6261 on Component 1, indicating its influential role in this principal component. The variable "1-person household" is notably associated with Component 1 as well, with a loading of -0.3353, which implies a significant but inverse relationship with this component. Furthermore, "government payment" displayed a substantial positive loading on Component 2, equating to 0.858, highlighting its primary role in explaining the variance within this component.

While these variables and their respective loadings offer a glimpse into the key factors influencing the various components, it is also critical to consider the eigenvalues, which represent the variance explained by each component. An eigenvalue threshold of 0.5 was applied to discern the significance of the components. Consequently, components with eigenvalues exceeding this threshold are deemed to significantly encapsulate the variance within the dataset, thereby meriting further exploration and interpretation. The comprehensive investigation of variables with substantial loadings in components, alongside considerations of respective

eigenvalues, facilitates a more nuanced understanding of the predominant patterns within the data, thereby informing targeted strategic approaches in the context of life insurance within Canadian households. This nuanced understanding could potentially translate into actionable insights for stakeholders within the insurance domain, assisting in formulating strategies that are optimally aligned with discerned patterns and trends.

The post-rotation phase of our analysis was particularly enlightening. Several data-driven narratives emerged:

The inaugural component was delineated by its robust affinity with demographic characteristics. Households typified as 'Couples with kid(s)' and metrics related to the 'Size of household' were dominant players. In a somewhat contrasting narrative, 'Single-person households' registered a negative loading on this component, suggesting varied insurance behaviours across household types.

The narrative surrounding the second component was multifaceted. Households that were predominantly reliant on governmental assistance as a primary income channel emerged as significant, registering a pronounced positive loading. Concurrently, income metrics displayed an inverse relationship with this component, a finding that resonated with broader socio-economic patterns. The implications were clear: higher income tiers bolster insurance allocations—a revelation that harmoniously aligns with existing literature. In contrast, households primarily sustained by government subsidies exhibited a tempered enthusiasm for life insurance investments, potentially reflecting socio-economic challenges.

While being of significant analytical interest, the tertiary component was distinctly shaped by metrics often associated with stability and future financial planning. 'Length of tenure' and

'Anticipated property value' were both indicative of a positive influence on life insurance outlays, suggesting that long-term planning and property valuation play a role in insurance decisions.

In exploring Canadian household life insurance expenditures, we employed a regression analysis that juxtaposed these expenditures, represented logarithmically, against the first 26 principal components. This approach yielded significant associations between each component and the dependent variable. Notably, while these associations were robust, the magnitudes of the corresponding coefficients were relatively subdued. This observation suggests a nuanced decision-making landscape in life insurance purchases within Canadian households. The incremental contributions of various determinants paint a complex and multifaceted picture of the factors influencing insurance decisions.

Central to our findings is the revelation that the 26 principal components collectively capture a substantial 94.78% of the total variance. This coverage highlights the depth and comprehensiveness of our PCA-centric approach.

Our analysis further delineates the relationships of these components with life insurance spending. Component 1, for instance, exhibits a positive correlation with life insurance expenditures. This is particularly pronounced in households with children, as evidenced by the strong positive loading for 'couple with kid(s).' Such households are inclined to allocate more resources to life insurance, presumably for ensuring financial security, especially concerning their children's future. In contrast, single-person households, as indicated by their negative loading, appear less inclined to significant life insurance investment, likely due to a reduced need stemming from fewer dependents.

Conversely, Component 2 displays an inverse relationship with life insurance spending.

Households predominantly reliant on government payments, as suggested by the strong positive

loading on 'government payment' within this component, tend to invest less in life insurance. This trend could be attributed to financial constraints or differing priorities in financial planning. In contrast, the negative loadings for groups such as 'one-person household' and 'couple with kid(s)' in this component suggest a higher propensity to invest in life insurance among households that are not primarily dependent on government support, potentially reflective of better financial stability or a different risk assessment framework.

Utilizing datasets spanning 2010 to 2017, a period marked by significant economic and socio-political shifts, we meticulously fine-tuned our model leveraging the PCA framework. This refined model was subsequently extrapolated to the 2019 household data, a year of interest for our study. Our prognostications, formulated using this robust methodology, pegged the average insurance expenditure for 2019 at \$1,263. However, when we compared the figures to the actual data, we found that they were slightly higher than our forecasts. The total came to \$1,381.

The model's projection was based on a rigorous methodology, but it did not perfectly align with the actual results. For the average 2019 Canadian household life insurance expenditure, our model's estimate stood at \$1,263 per insured household. Contrasting this with the empirically observed weighted mean of \$1,381, we discerned a variance of \$118 (Table 5). While our model's estimations hovered in proximity to actual figures, diverse elements—including external economic forces, regulatory shifts, and unforeseen variables—might have catalyzed this divergence. It is pivotal to contextualize the model's efficacy within this expansive spectrum of potential determinants. Such deviations, while noteworthy, are not uncommon in predictive analytics. Remarkably, a discrepancy of \$118, when juxtaposed against the vast landscape of household expenditures, underscores the model's commendable precision and its potential applicability in future research endeavours.

(insert Table 5 here)

Navigating through the intricate landscape of life insurance spending predictions and actual expenditures in 2019, our exploration across ten groups has unearthed pivotal insights and showcased the commendable strengths of the predictive model employed. Initially embarking on a methodical analysis of life insurance spending data, we sorted predictions of Canadian households' insurance spending and systematically categorized the weighted data into ten equitable groups. A calculation of the mean for each group, and subsequently, the actual insurance spending for each corresponding group (elaborated in Table 6 and Note: Figure 2), revealed a strikingly accurate predictive capability in several instances. Specifically, groups 1, 2, 3, 5, and 8 exhibited an impressively minimal divergence between predicted and actual life spending, maintaining a deviation constrained within a mere $\pm 5\%$. This undeniably attests to the model's robust predictive prowess in several demographic sectors. Even in instances where the model manifested discrepancies, such as the slightly larger variations observed in groups 4, 6, 7, 9 and 10, the invaluable insights derived from these outcomes provide a pathway for iterative refinement. The varied magnitudes and percentage changes across groups, while indicative of differential predictive accuracy, pave the way for targeted model optimization across diverse segments. Thus, while this analysis underscores the importance of continuous scrutiny and refinement of the predictive model, it simultaneously celebrates the model's successes, spotlighting its potential and laying a foundation upon which future predictive endeavours can build, ensuring a progressively more accurate alignment with actual spending trajectories in the life insurance domain. The model, therefore, not merely stands as a testament to the impactful confluence of data and predictive analytics but also as a catalyst for future advancements in predictive accuracy across a multitude of demographic landscapes.

(insert Note: Figure 2 and Table 6 here)

An exploration into the predictive model's accuracy reveals a particularly intriguing trend concerning higher insurance spending groups, where the mean exhibits a notably larger spread. This phenomenon could be intricately tied to the financial behaviours and risk mitigation strategies employed by higher-income families, often characterized by a multifaceted approach to safeguarding their financial stability. Notably, for these families, life insurance may constitute merely one facet of a broader financial strategy, intertwined with various other risk-hedging mechanisms, thereby introducing an additional layer of complexity and variability into their insurance spending patterns. This divergent approach towards insurance spending, particularly prevalent in higher-income demographics, inherently embeds a degree of unpredictability into the model, especially when compared to lower spending groups. Therefore, while the model showcases commendable predictive capabilities across numerous segments, it encounters heightened challenges when navigating through the intricacies of the higher insurance spending strata. This not only underscores the necessity for a nuanced and adaptive predictive model that is finely attuned to the multifarious financial landscapes across different income brackets but also opens avenues for exploring more tailored predictive models, which consider the multifactorial financial behaviours intrinsic to various demographic segments.

Private Health Insurance Predictive Model and Discussion

The predictive model of the Canadian households' private health insurance spending is as follows:

$$\hat{Y}_{PCA,health} = 2 \cdot (\beta_0 + \beta_1 \cdot (PC_1) + \beta_2 \cdot (PC_2) + \beta_3 \cdot (PC_3) + \dots + \beta_n \cdot (PC_n))$$

- β_0 is the intercept.
- β_i are the coefficients for the principal components PC_i

Our dataset revealed that out of 46,568 Canadian households, a significant number opted for private health insurance. Adjusting for the weighted data, this number is projected to represent approximately 6,126,880 households across Canada. It is noteworthy to mention that the overall Canadian household count approximates 14 million (Table 1).

To ascertain the rigour and uniformity of our analytical process, we applied a consistent methodological framework across disparate datasets. The emphasis was predominantly on Canadian households that had chosen private health insurance. It was fascinating to observe that the patterns discerned from the eigenvalue chart were congruent with patterns from our prior analyses on life insurance. The initial three components emerged as notably influential, encapsulating a significant fraction of the overall variance. From the fourth to the twenty-sixth component, the explanatory significance remained unwavering.

Following the implementation of the Principal Component Analysis (PCA) and its subsequent VariMax rotation, we discerned distinct regional tendencies. Specifically, households in Quebec (QBC) and Newfoundland and Labrador (NEW) exhibited a heightened affinity for private health insurance as opposed to other provinces. A nuanced exploration of the dataset underscored certain household attributes as critical determinants. For example, households identified under the category "couple with children" showcased a pronounced positive loading of 0.7169 within the inaugural component. Concurrently, the "size of the household" bore a compelling loading of 0.6216 within the same component. The second component underscored salient factors such as "tenure duration" (0.7944) and "anticipated residential resale value" (0.4944), both boasting robust positive loadings (Table 10). It is pivotal to note that while elements like "government subsidies" and "gross income" were prominently loaded in the third component, the latter manifested an inverse association with the dependent variable. This insinuates that a "gross

income" surge correlates positively with insurance outlay. However, households primarily sustained by government subsidies tend to exhibit a reticence towards procuring private health insurance. An exhaustive delineation of these revelations, accompanied by their respective loadings, is elucidated in Table 5.

(Insert Table 5, Table 10 here)

Our predictive framework, capitalizing on PCA for the curtailment of dimensionality, postulated the average 2019 expenditure of Canadian households on private health insurance to hover around \$1608. In contrast, the empirically observed average was pegged at \$1594, marking a marginal variance of \$14.

This minuscule discrepancy of \$14 between the anticipated and actualized values vouches for the predictive model's astuteness in approximating the expenditure trajectory of Canadian households vis-à-vis private health insurance. Given the convoluted dynamics and capricious tendencies characteristic of financial datasets, such a marginal divergence accentuates the robustness and precision of our PCA-augmented predictive paradigm.

Table 7 and

Note: Figure 4 provided data underscores a fascinating yet complex dynamic in spending prediction disparities among ten distinct groups. A conspicuous variance between anticipated and actual spending is noted, with certain clusters, notably Groups 1, 2, 4, 7, and 10, overshooting their financial forecasts, while a contrasting underestimation is exhibited by Groups 3, 5, 6, 8, and 9. Particularly, Group 1 overtly exceeded their predictions, overshooting by an assertive 11.45% or \$158.08, while Group 8 markedly undervalued their actual spending by a stark \$217.12 or -15.72%. Interestingly, while certain collectives like Group 6 displayed a minuscule disparity of \$8.77 or -0.63%, presenting a near-accurate forecast, others, such as Group 8,

highlighted a more substantial discord, hinting at challenges in prediction or unforeseen spending events. These divergences hint towards a multifaceted issue, potentially mandating a meticulous re-evaluation of the employed predictive analytics. The alternating pattern of over and underprediction across these groups does not follow a uniform trend (

Note: Figure 4, Table 7), either in terms of magnitude or directionality, suggesting that the inconsistencies may not be anchored in a singular bias towards overestimation or underestimation. Furthermore, while certain groups like 7 and 10 displayed analogous absolute differences, the proportional disparities provide additional depth, accentuating the necessity to gauge variations in both absolute and percentage metrics for a thorough financial analysis and consequently recalibrating predictive mechanisms to enhance future forecasting reliability.

(Insert Figure 4, Table 7 here)

Car Insurance Predictive Model and Discussion

The predictive model of Canadian households' car insurance spending is as follows:

$$\hat{Y}_{PCA,car} = \frac{1}{2} \cdot (\beta_0 + \beta_1 \cdot (PC_1) + \beta_2 \cdot (PC_2) + \beta_3 \cdot (PC_3) + \dots + \beta_n \cdot (PC_n))$$

- β_0 is the intercept.
- β_i are the coefficients for the principal components PC_i

In our comprehensive analysis of car insurance data using Principal Component Analysis (PCA), distinct patterns emerged warranting further scrutiny. Two components, both with eigenvalues surpassing 1, were especially pronounced, signifying their substantial influence on the dataset's variance. Interestingly, after the fourth component, there was a marked decrease in relevance, suggesting that the unaccounted variance in our model fluctuated between 10% and 30%.

Upon rotation, a wealth of insights surfaced. The first component revealed significant loadings tied to various household compositions: a loading of -0.3503 for single-person households,

0.6963 for couples with children, and 0.613 capturing the broader household size. Component 2 was characterized by strong loadings reflecting the duration of homeownership at 0.7921 and the anticipated property resale value at 0.5135. Component 3 was predominantly influenced by factors related to income and educational levels. Remarkably, the fourth component was almost singularly driven by urban considerations, highlighting its critical importance. (Table 11)

(Insert Table 11 here)

A key revelation from our research was the pronounced role of geographical factors on insurance spending. Households in regions such as British Columbia (BC), Manitoba (MA), Alberta (AB), and Newfoundland and Labrador (NEW) exhibited a stronger propensity to allocate more towards car insurance, diverging from trends in other provinces. This regional distinction underscores the criticality of understanding local dynamics when forecasting insurance expenditure patterns.

In analyzing Canadian car insurance spending, it's important to consider the potential reasons for the discrepancies between predicted and actual expenditures across provinces. This includes acknowledging the diverse nature of insurance providers—public insurers in provinces like British Columbia and Manitoba, private ones in Ontario and Alberta, and hybrids in Quebec and Saskatchewan. These regional policy differences significantly influence household insurance expenses. A comprehensive analysis reveals that provinces with public insurance, particularly BC and Manitoba, show a forward movement in component weights for car insurance spending, suggesting a consistent trend with the type of provincial insurer. This observation is key to understanding national insurance spending patterns.

However, car insurance was a different story. The model overshot the mark by a significant 19.07%. This indicates there might be other variables that could impact car insurance costs.

Zooming in on ten different household groups, a pattern emerged. The model consistently overestimated car insurance spending for all groups except the highest spenders. Overestimation ranged from 4.0% to 24.4%. This suggests a bias in the model, or perhaps missing variables, that particularly affect lower-spending households.

On the flip side, the model's performance for car insurance spending itself wasn't a complete miss. It predicted an average household expenditure of \$2,048 for car insurance in 2019. The actual average was \$1,720, meaning the model underestimated by \$328. While this is a modest variance, it highlights the inherent challenges of prediction and areas for improvement.

Delving into a critical evaluation of Note: Figure 6 and Table 8, which presents the predicted and actual mean car insurance spending across ten distinct groups within Canadian households, pronounced discrepancy surfaces, meriting a meticulous exploration and discussion about the accuracy and reliability of predictive models utilized. The data encompasses a variety of mean spending, spanning from \$591.71 to \$1,695.93 in predicted values and \$255.07 to \$1,719.81 in actual expenditures across groups 1 through 10, respectively. Notably, in 9 out of the 10 groups, the anticipated spending perceptibly overshoots the actual spending, with group 1 witnessing the most sizable difference of \$336.63, equating to a 24.3758% overestimation. Conversely, group 10 stands out as an anomaly, wherein the actual spending slightly surpasses the predicted by \$23.88, accounting for a -1.7292% variance.

(Insert Note: Figure 6, Table 8 here)

The prevalent discrepancies, particularly the overarching trend of overestimations across the majority of groups and the contrasting underestimation in Group 10 unfold a multi-layered discussion regarding the data, variables, and methodologies encapsulated within the predictive

models. Given the substantial dataset of 8,251 observations and a significant total population size of 10,395,761, the findings underline a pivotal need for refining predictive methodologies.

A thorough inquiry into whether the predictive models were aptly calibrated to accommodate potential caps on insurance spending due to vehicle values and whether they adeptly navigated the complexities and restrictions imposed by such value constraints is paramount. This exploration potentially hints at a requirement for the predictive models to more closely align with and represent the variables and characteristics inherent to each group, thereby optimizing the precision and reliability of future forecasts in car insurance spending across various segments of Canadian households. This approach would, in theory, contribute to enhancing the alignment between predicted and actual spending, driving more informed and strategic decision-making within the insurance sector and facilitating more accurate future financial planning and policymaking.

Chapter Four

Conclusion:

Our study concentrated on the complex elements that affect the process of making insurance decisions among Canadian households. By employing principal component analysis (PCA), we were able to refine our analysis and mitigate challenges such as multicollinearity, thereby revealing significant insights into the determinants of insurance decisions across a range of demographic groups, economic conditions, and regions.

The predictive model, developed using principal component analysis (PCA) and machine learning techniques, provided robust forecasts for household insurance spending in 2019. The model predicted the average expenditure for life insurance to be \$1,263, for private health insurance, \$1,608, and for car insurance, \$2,048. A comparison of the model's predictions with

actual spending revealed commendable accuracy for private health insurance, with a slight deviation of only \$14 from the actual average of \$1594. The prediction for life insurance was also relatively accurate, with an actual average of \$1,381, indicating a minor underestimation of \$118. However, the model exhibited a larger variance for car insurance, with an overestimation of actual spending by \$328.

The 2019 insurance spending data from Canadians demonstrated the efficacy of the prediction model. The model demonstrated particular efficacy in forecasting life and health insurance expenditures. The predictions were found to be in close alignment with reality, with deviations of -8.54% and 0.88%, respectively. This indicates that the model effectively identified the primary factors influencing expenditure on these types of insurance.

Our research makes a significant theoretical contribution to the field of insurance by advancing the use of predictive analytics. By integrating principal component analysis with machine learning, we demonstrated how high-dimensional data can be effectively reduced and analyzed to identify the most significant variables influencing insurance spending. This approach not only enhances the understanding of the factors driving household insurance expenditures but also establishes a precedent for future studies employing similar methodologies in other domains.

Practically, the insights from our research offer valuable applications for multiple stakeholders. Insurance companies can leverage the predictive model to develop more effective and personalized insurance products tailored to different demographics. Policymakers can use the findings to design policies that address gaps in insurance coverage, promoting competition and innovation in the sector. Consumers benefit by making more informed financial decisions, potentially leading to better coverage and cost savings, thus enhancing their financial security.

The model also aids financial advisors and investors in providing accurate guidance and making strategic investment choices. Additionally, the study's methodology and data serve as a foundation for further studies, fostering innovation and a deeper understanding of insurance and household finance. Overall, the paper significantly enhances decision-making processes, contributing to more effective insurance products, informed policies, empowered consumers, and a more robust and equitable insurance market.

A significant finding was the high degree of accuracy demonstrated by the predictive model for life insurance expenditures, with only a minimal discrepancy between the predicted and actual spending. However, the model demonstrated certain limitations, particularly in the projections for private health and car insurance, where discrepancies were identified, indicating potential areas for improvement. The results underscored the pivotal role of factors such as household type, size of households, income levels, length of tenure property valuations, and regional dynamics in influencing insurance behaviours.

It is of the utmost importance to continuously refine and calibrate predictive models in order to ensure that they remain attuned to evolving landscapes and intricate dynamics. This study highlights the critical importance of perpetual model refinement in predictive analytics to enhance precision and reliability in forecasting insurance expenditures. Future research should concentrate on integrating more sophisticated machine learning algorithms and investigating additional variables to further improve the model's predictive capacity and applicability across diverse scenarios.

In conclusion, our research offers substantial theoretical and practical contributions to the field of insurance analytics. It provides a robust framework for understanding and predicting household insurance spending. The predictive model developed in this study has significant

potential for application in real-world scenarios, enhancing decision-making processes for various stakeholders in the insurance industry.

Future direction:

Future research should delve deeper into the intricate factors shaping insurance spending across Canada by building on the initial insights from our principal component analysis (PCA). This exploration should consider shifting trends over time to provide valuable historical context and enhance predictive modelling through advanced machine learning algorithms and a broader range of variables, including macroeconomic indicators and societal trends. Additionally, qualitative research methods, such as surveys and interviews, can offer insights into the motivations and challenges influencing insurance spending decisions across different household types. A critical analysis of public policy impacts is also necessary to understand the complex relationship between policymaking and consumer behaviour. Furthermore, a comparative study with global insurance spending trends can help identify unique Canadian phenomena and assess the applicability of international strategies within the domestic context. By pursuing these avenues, we aim to create a comprehensive and accurate model that reflects the complex reality of the Canadian insurance landscape.

Bibliography:

- Browne, M. J., & Kim, K. (1993). An international analysis of life insurance demand. *Journal of Risk and Insurance*, 616–634.
- Bro, R., & Smilde, A. K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, 17(1), 16–33. <https://doi.org/10.1002/cem.773>
- Buccioli, A., & Miniaci, R. (2011). Household portfolios and implicit risk preference. *The Review of Economics and Statistics*, 93(4), 1235–1250.
https://doi.org/10.1162/REST_a_00138
- Cacace, M., & Schmid, A. (2008). The healthcare systems of the USA and Canada: Forever on divergent paths? *Social Policy & Administration*, 42(4), 396–417.
<https://doi.org/10.1111/j.1467-9515.2008.00611.x>
- Calvet, L. E., Campbell, J. Y., & Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, 115(5), 707–747. JSTOR.
<https://doi.org/10.1086/524204>
- Campbell, J. Y. (2006). Household finance. *The Journal of Finance*, 61(4), 1553–1604.
<https://doi.org/10.1111/j.1540-6261.2006.00883.x>
- Campbell, R. A. (1980). The demand for life insurance: An application of the economics of uncertainty. *The Journal of Finance*, 35(5), 1155–1172. <https://doi.org/10.2307/2327091>
- De Bromhead, A., & Borowiecki, K. (2016). Immigration and the demand for life insurance: Evidence from Canada, 1911. *European Review of Economic History*, 20, 147–175.
<https://doi.org/10.1093/ereh/hev022>

- Di Matteo, L., & Herbert Emery, J. C. (2002). Wealth and the demand for life insurance: Evidence from Ontario, 1892. *Explorations in Economic History*, 39(4), Article 4.
[https://doi.org/10.1016/S0014-4983\(02\)00004-9](https://doi.org/10.1016/S0014-4983(02)00004-9)
- Guiso, L., Haliassos, M., & Jappelli, T. (2003). Household stockholding in Europe: Where do we stand and where do we go? *Economic Policy*, 18(36), 123–170.
<https://doi.org/10.1111/1468-0327.00104>
- Hwang, T., & Gao, S. (2003). The determinants of the demand for life insurance in an emerging economy—the case of China. *Managerial Finance*.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151.
- Kanchan, B. D., & Kishor, M. M. (2016). Study of machine learning algorithms for special disease prediction using the principle of component analysis. 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 5–10. <https://doi.org/10.1109/ICGTSPICC.2016.7955260>
- Law, M. R., Daw, J. R., Cheng, L., & Morgan, S. G. (2013). Growth in private payments for health care by Canadian households. *Health Policy (Amsterdam, Netherlands)*, 110(2–3), Article 2–3. <https://doi.org/10.1016/j.healthpol.2013.01.014>
- Liu, Y., Zhang, Y., Chen, X., & Yang, Y. (2021). Superstition and farmers' life insurance spending. *Economics Letters*, p. 206, 109975. <https://doi.org/10.1016/j.econlet.2021.109975>
- Mapharing, M., Otuteye, E., & Radikoko, I. (2015). Determinants of demand for life insurance: The case of Canada. *Journal of Comparative International Management*, 18(2), 1–22.

Merkoulova, Y., & Veld, C. (2022). Why do individuals not participate in the stock market?

International Review of Financial Analysis, 83, 102292.

<https://doi.org/10.1016/j.irfa.2022.102292>

Nam, Y., & Hanna, S. D. (2019). The effects of risk aversion on life insurance ownership of single-parent households. *Applied Economics Letters*, 26(15), Article 15.

Outreville, J. F. (2013). The relationship between insurance and economic development: 85 empirical papers for a review of the literature. *Risk Management and Insurance Review*, 16(1), Article 1.

Tremblay, J., Lynch, J., & Dubreuil, G. (2010). Pilot survey results from the Canadian survey of household spending redesign—Joint Statistical Meetings of the American Statistical Association, Vancouver, British Columbia.

Wang, J., & Wang, J. (2015). Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks. *Neurocomputing*, 156, 68–78.

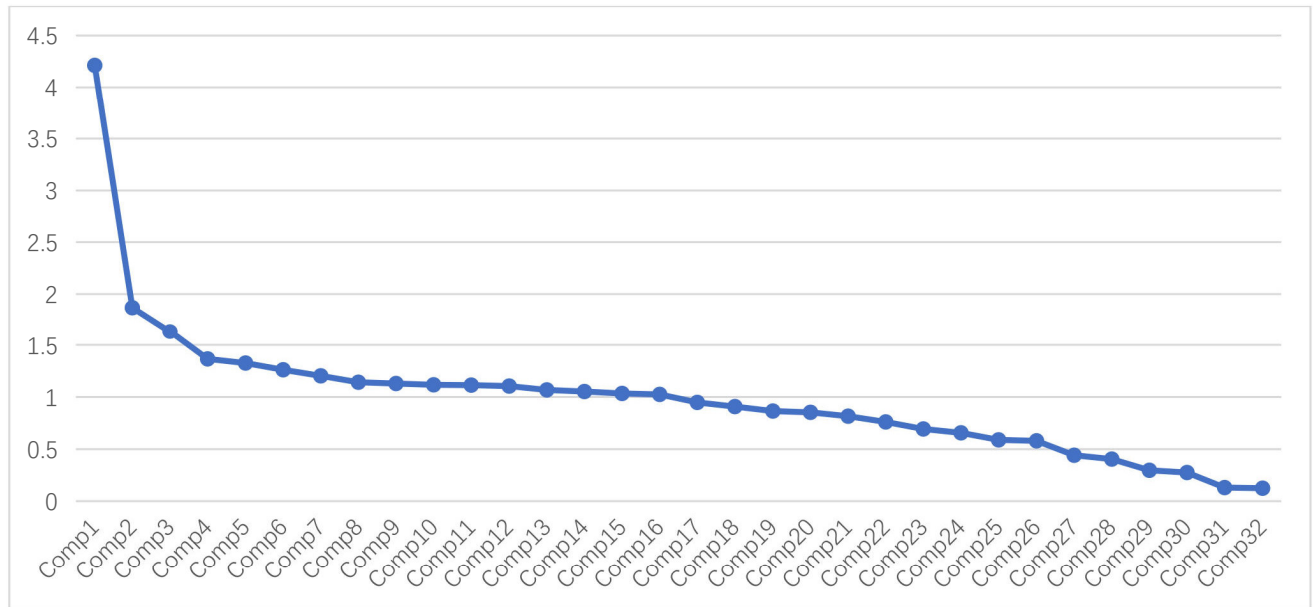
<https://doi.org/10.1016/j.neucom.2014.12.084>

Ye, Y., Pu, Y., & Xiong, A. (2022). The impact of digital finance on household participation in risky financial markets: Evidence-based study from China. *PLOS ONE*, 17(4), e0265606.

<https://doi.org/10.1371/journal.pone.0265606>

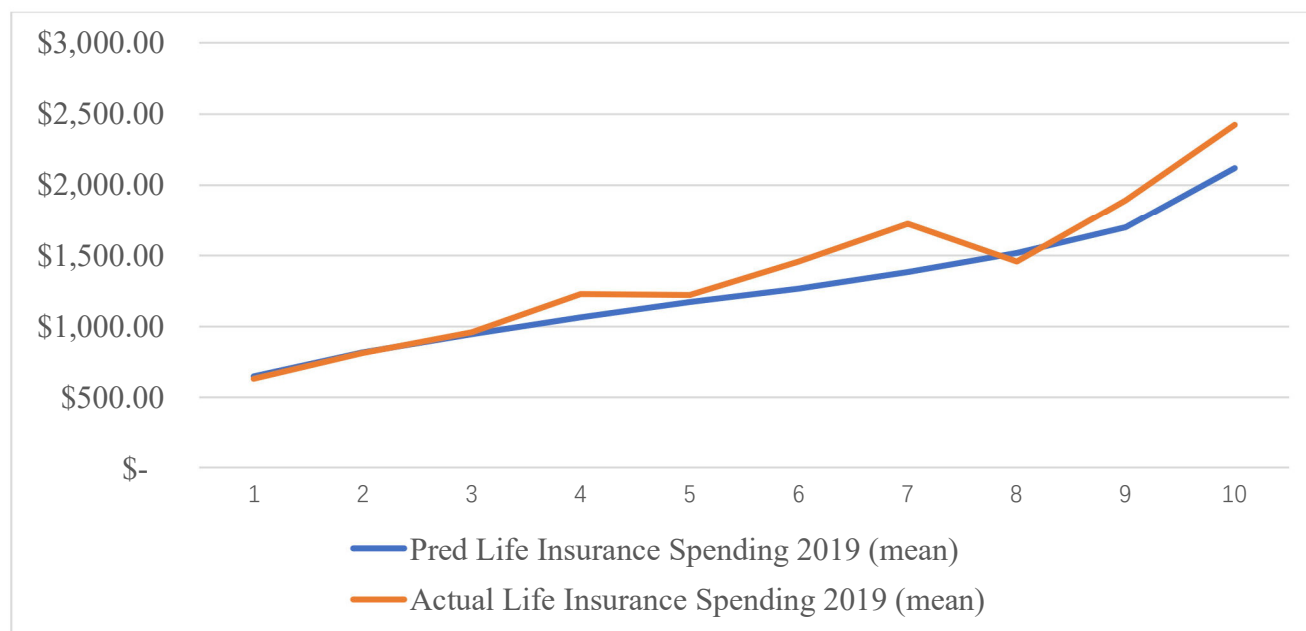
Appendix 1: Figures

Figure 1 Eigenvalue Distribution After PCA Analysis of Life Insurance Spending



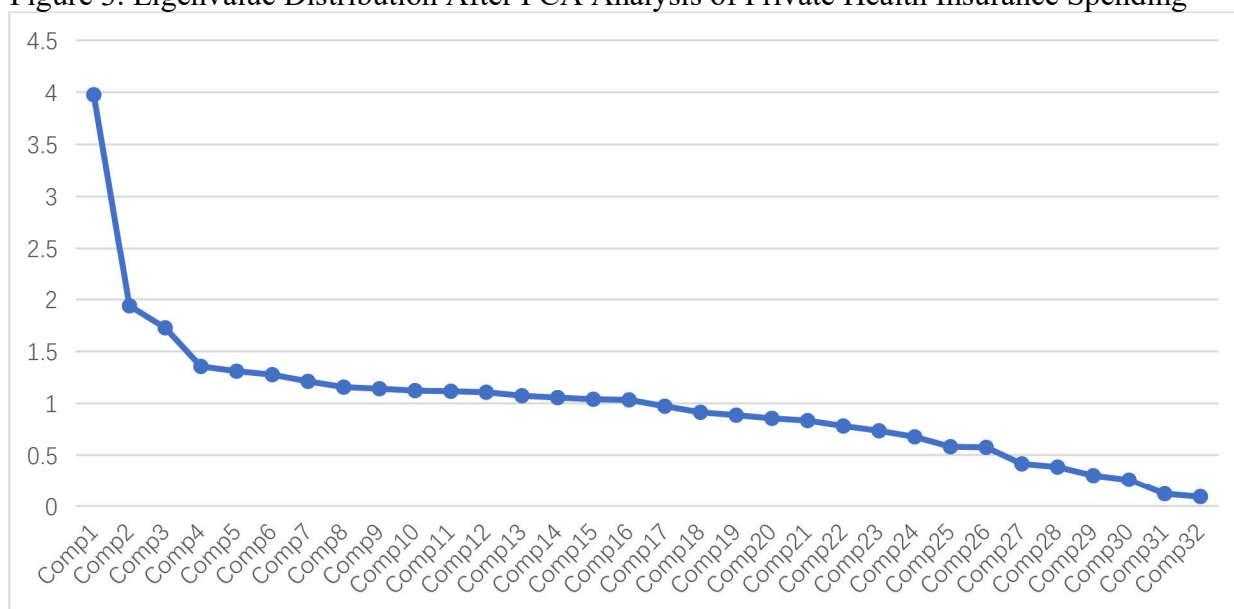
Note: Figure 1 is generated from Table 2, illustrates the eigenvalues of each component after performing PCA analysis on the independent variables. The slope from components 4 to 26 is gradual. A noticeable dip after component 26 suggests that subsequent components contribute even less to the variance.

Figure 2. Mean Predicted vs. Actual Life Insurance Spending Across 10 Groups in Canada, 2019



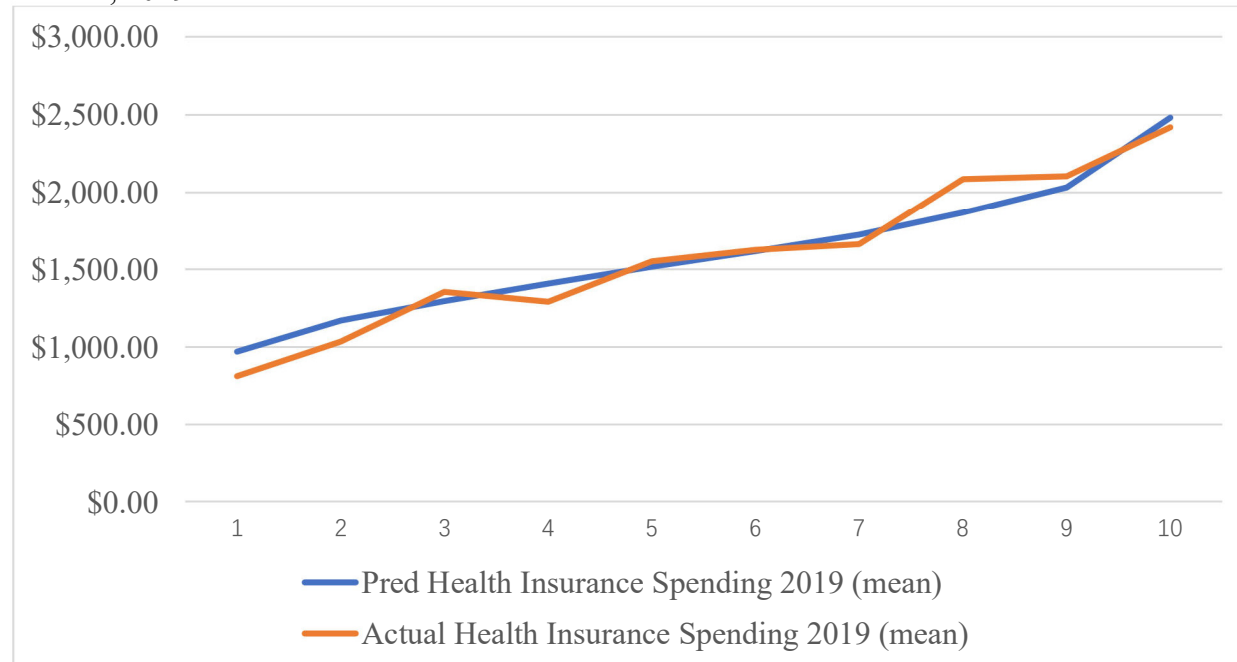
Note: Figure 2. generated from Table 6, shows that the predicted life insurance spending aligns with the actual spending. However, at higher spending groups, the predictions tend to be lower than the actual values, causing the spread to increase.

Figure 3. Eigenvalue Distribution After PCA Analysis of Private Health Insurance Spending



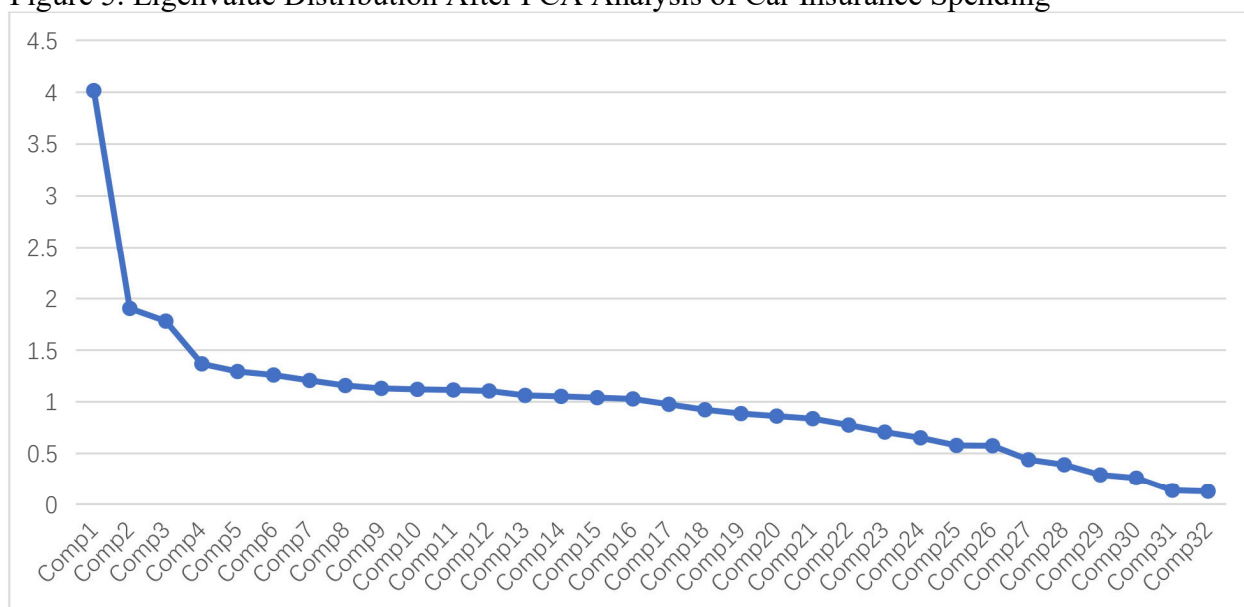
Note: Figure 3. This graph, generated from Table 3, illustrates the eigenvalues of each component after performing PCA analysis on the independent variables. The slope from components 4 to 26 is gradual. A noticeable dip after component 26 suggests that subsequent components contribute even less to the variance.

Figure 4. Mean Predicted vs. Actual Private Health Insurance Spending Across 10 Groups in Canada, 2019



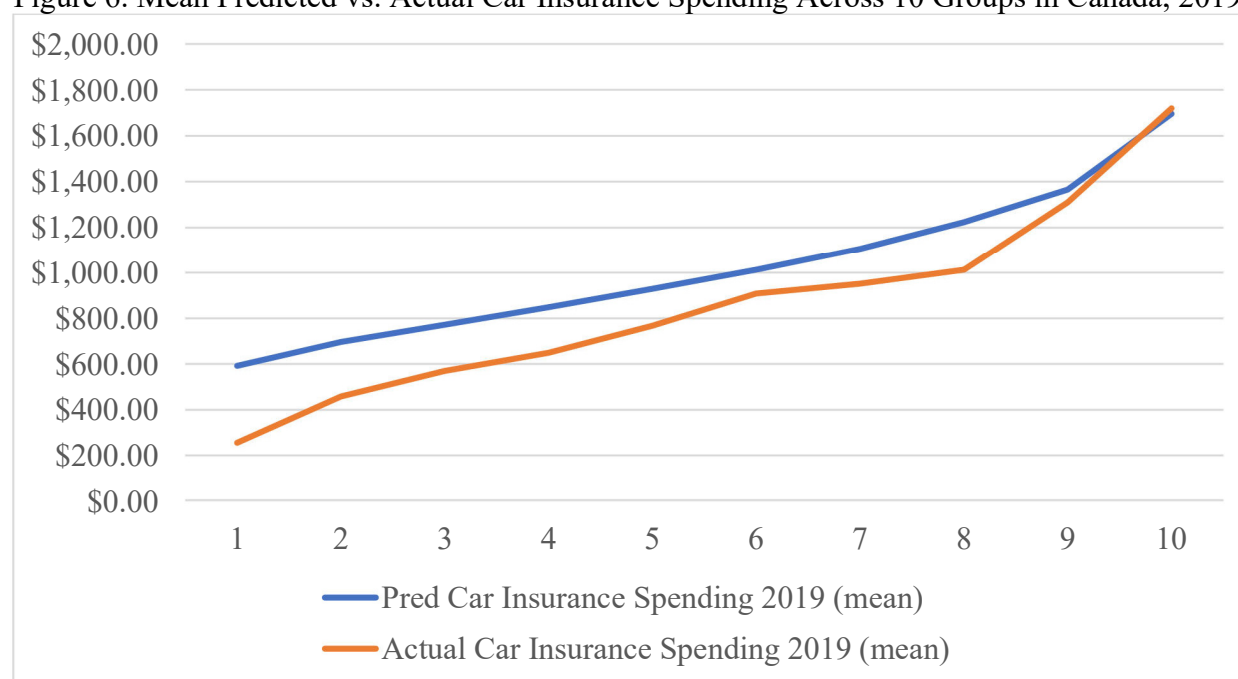
Note: Figure 4. The line graph compares mean predicted private health insurance spending to actual spending across ten groups in Canada for 2019. Both predicted (blue line) and actual (orange line) spending show an upward trend across the groups. In the initial groups (1 to 5), actual spending is slightly lower than predicted spending, indicating minor overestimation. In the higher groups (6 to 10), actual spending intersects with and occasionally exceeds predicted spending, suggesting improved accuracy in predictions. Overall, the graph demonstrates that while predictions are generally close to actual spending, there are slight discrepancies, particularly in the lower groups, with better alignment in higher-spending groups.

Figure 5. Eigenvalue Distribution After PCA Analysis of Car Insurance Spending



Note: Figure 5. is generated from Table 4, illustrates the eigenvalues of each component after performing PCA analysis on the independent variables. The slope from components 4 to 26 is gradual. A noticeable dip after component 26 suggests that subsequent components contribute even less to the variance.

Figure 6. Mean Predicted vs. Actual Car Insurance Spending Across 10 Groups in Canada, 2019



Note: Figure 6. The line graph compares mean predicted car insurance spending to actual spending across ten groups in Canada for 2019. It shows an upward trend for both predicted and actual spending as group numbers increase. Initially, predicted spending (blue line) is higher than actual spending (orange line) for most groups, indicating overestimation. However, this gap narrows towards the higher groups, where actual spending meets and eventually exceeds predicted spending around groups 9 and 10. This suggests that while predictions were generally higher, actual spending aligns more closely with predictions in higher-spending groups, highlighting discrepancies primarily in lower groups.

Appendix 2: Tables

Table 1. Number of Observations and Corresponding Weighted Population Estimates.

2010-2017			2019	
	Observations	weighted population	Observations	weighted population
life	37,610	5,610,380	5,315	6,822,934
health	46,568	6,126,880	6,425	7,610,624
car	75,451	11,153,340	8,251	10,395,761

Note: Table 1 shows the number of observations and corresponding weighted population estimates for life, health, and car insurance from 2010-2017 and in 2019. While the number of observations for life and car insurance decreased in 2019, the weighted population estimates for life and health insurance increased during this period.

Table 2. Eigenvalue Distribution After PCA Analysis of Life Insurance Spending

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	4.212	2.345	0.132	0.132
Comp2	1.867	0.231	0.058	0.190
Comp3	1.635	0.268	0.051	0.241
Comp4	1.367	0.040	0.043	0.284
Comp5	1.327	0.065	0.042	0.325
Comp6	1.261	0.057	0.039	0.365
Comp7	1.204	0.064	0.038	0.402
Comp8	1.141	0.010	0.036	0.438
Comp9	1.130	0.012	0.035	0.473
Comp10	1.119	0.005	0.035	0.508
Comp11	1.113	0.010	0.035	0.543
Comp12	1.104	0.035	0.035	0.578
Comp13	1.068	0.016	0.033	0.611
Comp14	1.053	0.017	0.033	0.644
Comp15	1.035	0.012	0.032	0.676
Comp16	1.024	0.074	0.032	0.708
Comp17	0.949	0.041	0.030	0.738
Comp18	0.908	0.043	0.028	0.766
Comp19	0.865	0.013	0.027	0.793
Comp20	0.853	0.037	0.027	0.820
Comp21	0.815	0.056	0.026	0.845
Comp22	0.759	0.066	0.024	0.869
Comp23	0.693	0.037	0.022	0.891
Comp24	0.656	0.066	0.021	0.911
Comp25	0.590	0.010	0.018	0.930
Comp26	0.580	0.138	0.018	0.948
Comp27	0.442	0.037	0.014	0.962
Comp28	0.404	0.107	0.013	0.974
Comp29	0.298	0.024	0.009	0.984
Comp30	0.274	0.144	0.009	0.992
Comp31	0.129	0.005	0.004	0.996
Comp32	0.125	.	0.004	1.000

Note: Table 2 shows the eigenvalue distribution from PCA analysis of life insurance spending, with the first component explaining 13.2% of the total variance and the first three components cumulatively explaining 24.1%. As the component number increases, their individual contributions to variance decrease, indicating that only a few components account for most of the variance.

Table 3. Eigenvalue Distribution After PCA Analysis of Private Health Insurance Spending

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	3.978	2.037	0.124	0.124
Comp2	1.940	0.212	0.061	0.185
Comp3	1.728	0.374	0.054	0.239
Comp4	1.354	0.046	0.042	0.281
Comp5	1.308	0.033	0.041	0.322
Comp6	1.275	0.065	0.040	0.362
Comp7	1.210	0.055	0.038	0.400
Comp8	1.155	0.014	0.036	0.436
Comp9	1.140	0.019	0.036	0.472
Comp10	1.121	0.007	0.035	0.507
Comp11	1.114	0.009	0.035	0.541
Comp12	1.106	0.033	0.035	0.576
Comp13	1.073	0.019	0.034	0.609
Comp14	1.054	0.014	0.033	0.642
Comp15	1.040	0.009	0.033	0.675
Comp16	1.031	0.060	0.032	0.707
Comp17	0.971	0.058	0.030	0.737
Comp18	0.913	0.028	0.029	0.766
Comp19	0.885	0.031	0.028	0.794
Comp20	0.855	0.022	0.027	0.820
Comp21	0.833	0.051	0.026	0.846
Comp22	0.782	0.047	0.024	0.871
Comp23	0.735	0.058	0.023	0.894
Comp24	0.677	0.096	0.021	0.915
Comp25	0.580	0.007	0.018	0.933
Comp26	0.574	0.158	0.018	0.951
Comp27	0.415	0.030	0.013	0.964
Comp28	0.385	0.084	0.012	0.976
Comp29	0.301	0.043	0.009	0.985
Comp30	0.258	0.138	0.008	0.993
Comp31	0.120	0.029	0.004	0.997
Comp32	0.091	.	0.003	1.000

Note: Table 3 provides the eigenvalue distribution from a Principal Component Analysis (PCA) of health insurance spending, highlighting how each component contributes to explaining the total variance in the data. It shows that a few principal components account for a significant portion of the variance, while the contributions of subsequent components gradually diminish.

Table 4. Eigenvalue Distribution After PCA Analysis of Car Insurance Spending

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	4.013	2.109	0.125	0.125
Comp2	1.904	0.125	0.060	0.185
Comp3	1.778	0.412	0.056	0.241
Comp4	1.367	0.074	0.043	0.283
Comp5	1.293	0.036	0.040	0.324
Comp6	1.257	0.052	0.039	0.363
Comp7	1.205	0.048	0.038	0.401
Comp8	1.157	0.030	0.036	0.437
Comp9	1.128	0.007	0.035	0.472
Comp10	1.120	0.009	0.035	0.507
Comp11	1.112	0.007	0.035	0.542
Comp12	1.105	0.045	0.035	0.576
Comp13	1.060	0.008	0.033	0.609
Comp14	1.052	0.012	0.033	0.642
Comp15	1.040	0.012	0.033	0.675
Comp16	1.028	0.054	0.032	0.707
Comp17	0.974	0.052	0.030	0.737
Comp18	0.921	0.037	0.029	0.766
Comp19	0.884	0.023	0.028	0.794
Comp20	0.860	0.026	0.027	0.821
Comp21	0.834	0.059	0.026	0.847
Comp22	0.775	0.069	0.024	0.871
Comp23	0.706	0.056	0.022	0.893
Comp24	0.649	0.075	0.020	0.913
Comp25	0.575	0.003	0.018	0.931
Comp26	0.572	0.135	0.018	0.949
Comp27	0.437	0.050	0.014	0.963
Comp28	0.387	0.098	0.012	0.975
Comp29	0.289	0.031	0.009	0.984
Comp30	0.258	0.124	0.008	0.992
Comp31	0.134	0.008	0.004	0.996
Comp32	0.125	.	0.004	1.000

Note: Table 4 provides the eigenvalue distribution from a Principal Component Analysis (PCA) of car insurance spending, highlighting how each component contributes to explaining the total variance in the data. It shows that a few principal components account for a significant portion of the variance, while the contributions of subsequent components gradually diminish.

Table 5. 2019 Canadian Households Insurance Spending

Insurance Type	Predicted Spending (mean)	Actual Spending (mean)	Difference	Percentage
Life Insurance	1263	1381	-118	-8.54%
Health Insurance	1608	1594	14	0.88%
Car Insurance	2048	1720	328	19.07%

Note: Table 5. provides an analysis of Canadian households' insurance spending in 2019, comparing predicted spending to actual spending across different types of insurance.

Table 6. Mean Life Insurance Spending for Canadian Households: Averages Across Ten Distinct Groups

groups	Pred Life Insurance Spending 2019 (mean)	Actual Life Insurance Spending 2019 (mean)	difference in mean	Percent change of difference
1	\$648.16	\$631.44	\$16.71	1.2%
2	\$816.46	\$811.79	\$4.67	0.3%
3	\$945.13	\$958.36	-\$13.24	-1.0%
4	\$1,063.37	\$1,227.18	-\$163.81	-11.9%
5	\$1,172.85	\$1,221.17	-\$48.32	-3.5%
6	\$1,266.39	\$1,456.21	-\$189.82	-13.7%
7	\$1,382.58	\$1,724.51	-\$341.94	-24.8%
8	\$1,518.86	\$1,457.97	\$60.89	4.4%
9	\$1,699.46	\$1,891.80	-\$192.35	-13.9%
10	\$2,120.77	\$2,424.50	-\$303.74	-22.0%
Number of obs = 5,315				
Population size = 6,822,934				

Note: Table 6. presents a comparative analysis of predicted and actual mean life insurance spending for Canadian households across ten distinct groups for the year 2019.

Table 7. Mean Health Insurance Spending for Canadian Households:
Averages cross Ten Distinct Groups

groups	Pred Health Insurance Spending 2019 (mean)	Actual Health Insurance Spending 2019 (mean)	difference in mean	Percent change of difference
1	\$970.06	\$811.98	\$158.08	11.4%
2	\$1,170.10	\$1,034.61	\$135.49	9.8%
3	\$1,295.81	\$1,354.13	-\$58.31	-4.2%
4	\$1,408.65	\$1,291.86	\$116.79	8.5%
5	\$1,516.55	\$1,551.65	-\$35.11	-2.5%
6	\$1,618.15	\$1,626.92	-\$8.77	-0.6%
7	\$1,725.50	\$1,663.46	\$62.04	4.5%
8	\$1,868.42	\$2,085.54	-\$217.12	-15.7%
9	\$2,032.18	\$2,104.98	-\$72.80	-5.3%
10	\$2,481.41	\$2,419.36	\$62.05	4.5%
Number of obs =		6,425		
Population size =		7,610,624		

Note: Table 7 provides an analysis of mean health insurance spending for Canadian households in 2019, comparing predicted and actual spending across ten distinct groups.

Table 8. Mean Car Insurance Spending for Canadian Households:
Averages Across Ten Distinct Groups

groups	Pred Car Insurance Spending 2019 (mean)	Actual Car Insurance Spending 2019 (mean)	difference in mean	Percent change of difference
1	\$591.71	\$255.07	\$336.63	24.4%
2	\$696.04	\$458.08	\$237.97	17.2%
3	\$771.25	\$568.97	\$202.28	14.6%
4	\$847.95	\$648.22	\$199.73	14.5%
5	\$929.57	\$767.56	\$162.01	11.7%
6	\$1,012.16	\$907.78	\$104.38	7.6%
7	\$1,106.72	\$951.29	\$155.42	11.3%
8	\$1,222.83	\$1,012.88	\$209.95	15.2%
9	\$1,365.04	\$1,310.05	\$54.98	4.0%
10	\$1,695.93	\$1,719.81	-\$23.88	-1.7%
Number of obs =		8,251		
Population size =		10,395,761		

Note: Table 8 provides an analysis of mean car insurance spending for Canadian households in 2019, comparing predicted spending to actual spending across ten distinct groups.

Table 9. Canadian Household Data: Post-VariMax Rotation Analysis for Life Insurance Holders

lg_EP007_raj Coef.	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12	Comp13
Variable	0.0923	-0.0397	0.1261	-0.0015	0.0137	0.0045	0.0310	0.0015	-0.0048	0.0710	0.0653	-0.0045	0.0444
Prv_BC	-0.004	0.005	-0.015	0.000	0.057	0.053	0.060	-0.088	0.053	-0.005	-0.030	-0.035	0.914
Prv_PEI	-0.000	-0.007	0.007	-0.004	0.071	0.069	0.066	-0.011	0.063	0.003	0.021	0.941	-0.025
Prv_NS	-0.002	-0.005	0.005	-0.001	0.097	0.093	-0.908	-0.041	0.086	0.002	0.013	-0.076	-0.055
Prv_NB	-0.002	-0.011	0.009	-0.005	-0.876	0.118	0.115	-0.039	0.108	0.003	0.026	-0.097	-0.057
Prv_QBC	-0.013	0.014	-0.013	0.029	0.305	0.260	0.282	-0.437	0.234	-0.008	-0.062	-0.162	-0.345
Prv_MA	-0.004	0.004	-0.011	-0.002	0.067	0.064	0.067	-0.067	0.063	-0.005	-0.016	-0.047	-0.071
Prv_SA	-0.004	0.002	-0.008	-0.011	0.086	0.084	0.081	-0.031	-0.920	-0.003	0.007	-0.069	-0.044
Prv_AB	-0.005	0.017	-0.022	0.007	0.066	0.059	0.070	0.864	0.060	-0.007	-0.052	-0.034	-0.124
Prv_NEW	-0.002	-0.002	0.004	-0.005	0.104	-0.899	0.097	-0.030	0.093	0.003	0.019	-0.084	-0.047
1 Person Household	-0.335	-0.094	-0.134	0.059	0.022	0.011	0.014	0.013	0.016	-0.074	-0.028	-0.009	-0.012
Couple With Kid(S)	0.693	-0.034	-0.042	0.010	0.004	0.000	0.003	-0.001	0.013	-0.244	-0.021	-0.002	-0.005
Couple With Other	-0.022	0.001	-0.017	0.003	-0.002	-0.002	-0.001	-0.005	0.003	0.929	-0.016	0.002	-0.004
Lone Parent	-0.040	-0.023	-0.020	0.011	0.003	0.000	0.002	0.001	0.002	-0.042	-0.006	-0.001	-0.004
Other HH Type	-0.040	-0.005	-0.030	0.010	0.000	-0.001	0.000	-0.003	0.004	-0.046	-0.019	0.002	-0.006
Own With Mortgage	0.015	-0.016	-0.129	0.846	0.003	0.004	0.000	-0.006	0.023	0.006	-0.022	-0.003	-0.011
Miscellaneous Income Source	-0.001	0.001	0.011	0.012	0.001	0.001	0.001	0.003	-0.001	-0.001	0.005	-0.002	0.004
Self Employee	0.002	-0.015	0.004	-0.006	0.004	0.002	0.003	0.005	0.002	0.002	0.011	-0.004	0.002
Investment Income	0.000	-0.006	-0.003	-0.008	0.002	0.002	0.001	0.005	0.002	-0.001	0.005	-0.002	0.004
Government Payment	0.019	0.858	0.062	-0.019	0.015	-0.002	0.006	0.020	-0.011	0.018	0.081	-0.010	0.005
Other Major Source Income	-0.003	-0.034	0.003	-0.012	0.010	0.006	0.006	0.012	0.005	-0.003	0.022	-0.009	0.005
Suitable Place To Live	0.010	-0.002	0.015	0.002	-0.004	-0.005	-0.003	-0.007	-0.006	-0.005	0.004	0.004	-0.004
Have Internet	-0.003	-0.011	-0.002	-0.003	0.001	0.000	0.001	0.003	0.000	0.000	0.005	-0.001	0.001
Size of Household	0.626	0.038	-0.017	0.022	0.003	0.008	0.004	-0.006	-0.003	0.242	-0.014	0.000	-0.007
Length of Tenure	-0.022	0.028	0.812	-0.145	-0.011	-0.004	-0.008	-0.033	0.024	-0.025	-0.081	0.011	-0.030
House Selling Value	-0.022	0.028	0.498	0.501	0.005	0.002	0.008	0.038	-0.043	-0.006	0.072	-0.006	0.042
Higher Level of Edu	-0.012	0.036	-0.040	-0.004	-0.019	-0.014	-0.013	-0.033	-0.006	-0.016	0.949	0.019	-0.024
Urban Size	0.014	-0.039	0.058	-0.059	0.303	0.280	0.220	0.198	0.227	0.025	0.176	-0.236	0.100
Building Age	0.005	-0.014	0.023	0.010	0.009	0.007	0.007	0.008	0.002	0.007	0.021	-0.008	0.005
Income	0.095	-0.495	0.205	-0.044	0.023	-0.019	0.006	0.030	-0.041	0.081	0.194	-0.011	0.000
Rec Car Insurance	-0.005	0.000	-0.007	-0.008	0.013	0.014	0.009	0.021	0.013	-0.004	0.010	-0.014	0.016
Man Cloth Spending	-0.023	0.000	-0.020	0.004	0.002	0.002	0.001	0.005	0.004	-0.020	-0.010	-0.002	0.002
Female Cloth Spending	-0.022	-0.003	-0.020	0.000	0.005	0.005	0.003	0.009	0.007	-0.017	-0.005	-0.005	0.005

Table 9. Canadian Household Data: Post-VariMax Rotation Analysis for Life Insurance Holders (continued)

lg EP007_raj Coef.	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19	Comp20	Comp21	Comp22	Comp23	Comp24	Comp25	Comp26
Prv_BC	0.004	0.0399	0.0112	0.0485	-0.0190	0.0163	0.0310	-0.0600	-0.0097	-0.0136	0.0366	0.0176	0.0530
Prv_PEI	-0.002	-0.041	-0.006	0.000	0.000	0.002	-0.017	0.006	-0.006	0.007	0.003	0.001	0.006
Prv_NS	-0.001	-0.070	-0.004	-0.001	-0.001	0.001	-0.011	-0.010	0.005	-0.011	-0.005	-0.001	-0.003
Prv_NB	-0.002	-0.081	-0.007	-0.002	-0.002	0.002	-0.020	-0.015	0.003	-0.008	-0.004	-0.001	-0.001
Prv_QBC	0.021	-0.310	0.032	-0.003	-0.005	0.002	-0.061	-0.025	0.006	-0.015	-0.007	-0.002	-0.004
Prv_MA	0.001	0.929	0.002	-0.004	-0.005	0.009	0.009	-0.002	0.003	0.003	0.001	0.001	0.003
Prv_SA	-0.005	-0.061	-0.009	-0.009	-0.003	-0.017	-0.017	0.007	-0.007	0.001	-0.002	0.000	-0.003
Prv_AB	0.009	-0.104	0.018	0.000	-0.006	0.040	0.040	-0.013	0.020	0.005	0.009	0.005	0.011
Prv_NEW	-0.003	-0.068	-0.007	0.001	0.002	-0.020	-0.011	0.007	-0.009	-0.002	-0.004	0.000	-0.003
1 Person Household	-0.194	-0.013	-0.196	-0.308	-0.283	-0.051	0.016	0.133	-0.097	0.001	-0.013	-0.062	-0.011
Couple With Kid(S)	-0.072	-0.012	-0.066	-0.156	-0.160	-0.018	0.012	-0.024	-0.040	-0.004	-0.005	-0.009	-0.010
Couple With Other	-0.023	-0.005	-0.020	-0.042	-0.048	-0.005	0.008	-0.008	-0.005	-0.001	0.002	0.000	-0.002
Lone Parent	-0.044	-0.004	-0.043	0.931	-0.066	-0.010	0.003	0.020	-0.026	-0.001	-0.005	-0.014	-0.004
Other HH Type	-0.045	-0.005	-0.043	-0.063	0.934	-0.011	0.010	0.020	-0.014	0.000	0.002	-0.010	-0.002
Own With Mortgage	0.004	-0.012	-0.006	0.016	0.012	-0.018	0.032	0.010	-0.030	0.035	-0.015	-0.008	-0.022
Miscellaneous Income	0.001	0.002	0.001	-0.001	0.000	0.001	-0.005	-0.002	0.002	0.996	0.001	0.001	0.002
Self Employee	0.001	0.001	0.000	-0.003	0.002	-0.002	-0.005	-0.001	-0.012	0.001	0.994	-0.004	-0.003
Investment Income	-0.001	0.002	-0.003	-0.003	-0.001	-0.004	0.000	0.002	-0.007	0.002	-0.003	-0.002	0.997
Government Payment	0.014	0.007	0.006	-0.030	0.010	0.009	-0.038	-0.012	-0.078	0.001	-0.039	-0.026	-0.015
Miscellaneous Income Source	-0.007	0.002	-0.010	-0.020	-0.010	-0.007	-0.008	0.004	0.971	0.002	-0.013	-0.010	-0.008
Suitable Place To Live	0.017	-0.001	0.018	0.014	0.015	0.011	-0.005	0.982	0.004	-0.002	-0.001	0.004	0.002
Have Internet	-0.005	0.001	-0.006	-0.010	-0.007	-0.002	-0.002	0.004	-0.009	0.001	-0.004	0.996	-0.002
Size of Household	-0.009	0.002	-0.013	0.061	0.061	-0.003	0.005	0.091	0.027	0.003	0.012	-0.011	0.012
Length of Tenure	-0.035	-0.025	-0.039	-0.020	-0.041	-0.020	0.038	0.034	-0.007	0.038	0.003	-0.006	-0.015
House Selling Value	0.003	0.035	0.021	-0.017	-0.006	0.027	-0.075	-0.028	0.056	-0.074	0.026	0.018	0.041
Higher Level of Edu	-0.011	-0.013	-0.006	-0.004	-0.018	0.011	0.025	0.005	0.025	0.006	0.014	0.006	0.006
Urban Size	-0.023	-0.019	-0.056	0.018	0.029	-0.155	-0.053	0.047	-0.068	-0.012	-0.029	-0.008	-0.034
Building Age	0.006	0.002	0.005	0.002	0.008	0.000	0.987	-0.005	-0.007	-0.005	-0.005	-0.002	0.000
Income	0.068	0.015	0.054	-0.024	0.073	0.045	-0.092	-0.047	-0.164	0.002	-0.086	-0.051	-0.028
Rec Car Insurance	-0.011	0.005	-0.014	-0.007	-0.008	0.980	0.007	0.011	-0.008	0.000	-0.002	-0.002	-0.005
Man Cloth Spending	0.972	0.000	-0.028	-0.034	-0.037	-0.012	0.007	0.018	-0.008	0.001	0.001	-0.006	-0.002
Female Cloth Spending	-0.028	0.001	0.971	-0.034	-0.035	-0.015	0.006	0.019	-0.011	0.001	0.000	-0.007	-0.003

Note: Table 9. includes various variables and their corresponding coefficients across ten components (Comp1 to Comp10). The first three components (Comp1, Comp2, Comp3) from the Post-VariMax rotation analysis highlight critical factors influencing life insurance holdings among Canadian households. Component 1 shows a positive relationship with life insurance spending, indicating that households with couples and children, as well as larger household sizes, tend to spend more on life insurance. Component 2 has a negative relationship with life insurance spending, suggesting that households receiving government payments, typically with lower incomes, tend to spend less on life insurance. Component 3 also exhibits a positive relationship with life insurance spending.

Table 10. Canadian Household Spending on Private Health Insurance: A Post-VariMax Rotation Analysis

Variable	lg EP007_raj Coef.	0.04173	0.0908	-0.1001	-0.0078	-0.0347	-0.0017	0.08089	0.03466	-0.0784	0.04445	0.09534	0.04651	0.00119
Prv_BC	-0.002	-0.012	0.002	0.000	0.060	0.057	0.062	0.065	0.065	-0.083	0.002	-0.027	0.004	-0.003
Prv_PEI	0.000	0.006	-0.005	-0.004	0.073	0.071	0.068	0.069	0.069	-0.009	-0.003	0.019	-0.005	0.002
Prv_NS	-0.001	0.005	-0.004	-0.002	0.102	0.099	0.097	-0.901	-0.901	-0.039	-0.002	0.013	-0.003	0.002
Prv_NB	0.000	0.009	-0.009	-0.005	-0.881	0.115	0.111	0.113	0.113	-0.030	-0.003	0.024	-0.006	0.003
Prv_QBC	-0.002	-0.008	0.003	0.027	0.273	0.247	0.254	0.277	0.277	-0.494	0.016	-0.060	0.022	-0.004
Prv_MA	-0.003	-0.011	0.004	0.000	0.072	0.070	0.074	0.075	0.075	-0.068	-0.002	-0.018	-0.001	-0.005
Prv_SA	-0.003	-0.012	0.006	-0.012	0.099	0.097	-0.901	0.096	0.096	-0.034	-0.008	0.004	-0.009	-0.004
Prv_AB	-0.002	-0.019	0.011	0.010	0.080	0.075	0.084	0.090	0.090	0.836	0.008	-0.055	0.013	-0.006
Prv_NEW	0.002	0.007	-0.002	-0.005	0.106	-0.896	0.100	0.102	0.102	-0.026	0.000	0.020	-0.003	0.005
1 Person Household	-0.293	-0.122	-0.083	0.069	0.024	-0.003	0.019	0.016	0.016	0.011	-0.237	-0.034	-0.224	-0.065
Couple With Kid(S)	0.717	-0.036	-0.025	0.016	0.002	-0.006	0.013	0.000	0.000	0.000	-0.079	-0.020	-0.070	-0.213
Couple With Other	-0.026	-0.016	0.003	0.004	-0.002	-0.004	0.004	-0.002	-0.002	-0.003	-0.024	-0.015	-0.021	0.940
Lone Parent	-0.038	-0.015	-0.019	0.012	0.003	-0.002	0.003	0.002	0.002	0.002	-0.046	-0.005	-0.044	-0.037
Other HH Type	-0.042	-0.028	-0.005	0.013	0.000	-0.004	0.005	0.000	0.000	-0.001	-0.053	-0.018	-0.049	-0.043
Own With Mortgage	0.020	-0.123	-0.025	0.845	0.001	0.002	0.022	0.001	0.001	-0.003	0.005	-0.016	-0.003	0.008
Miscellaneous Income Source	-0.001	0.009	0.001	0.010	0.002	0.002	-0.001	0.002	0.002	0.003	0.001	0.005	0.001	0.000
Self Employee	0.002	0.005	-0.015	-0.006	0.004	0.002	0.001	0.003	0.003	0.003	0.001	0.011	0.001	0.002
Investment Income	0.001	-0.002	-0.008	-0.009	0.002	0.001	0.001	0.001	0.001	0.004	0.000	0.005	-0.001	0.000
Government Payment	0.014	0.076	0.865	-0.022	0.014	0.003	-0.016	0.007	0.007	0.015	0.012	0.078	0.009	0.018
Other Major Source Income	-0.001	0.014	-0.046	-0.015	0.010	0.005	0.001	0.006	0.006	0.008	-0.004	0.029	-0.005	0.001
Suitable Place To Live	0.008	0.014	-0.003	0.001	-0.002	-0.001	-0.004	-0.001	-0.001	-0.005	0.017	0.006	0.017	-0.004
Have Internet	-0.002	0.001	-0.010	-0.002	0.002	0.000	0.000	0.001	0.001	0.002	-0.004	0.005	-0.004	0.001
Size of Household	0.622	-0.006	0.034	0.018	0.001	0.005	-0.007	0.003	0.003	-0.001	0.005	-0.010	0.000	0.235
Length of Tenure	-0.020	0.794	0.042	-0.145	-0.014	-0.010	0.030	-0.009	-0.009	-0.029	-0.044	-0.090	-0.047	-0.026
House Selling Value	-0.032	0.494	0.054	0.501	0.009	0.009	-0.041	0.007	0.007	0.031	0.008	0.058	0.022	-0.009
Higher Level of Edu	-0.011	-0.042	0.036	-0.002	-0.017	-0.014	-0.002	-0.011	-0.011	-0.029	-0.014	0.950	-0.011	-0.015
Urban Size	0.010	0.042	-0.013	-0.058	0.308	0.289	0.254	0.236	0.236	0.192	-0.020	0.154	-0.036	0.018
Building Age	0.005	0.026	-0.017	0.010	0.007	0.006	-0.002	0.006	0.006	0.005	0.022	0.022	0.009	0.008
Income	0.077	0.281	-0.483	-0.050	0.029	0.002	-0.052	0.012	0.012	0.032	0.066	0.206	0.057	0.073
Rec Car Insurance	-0.006	-0.009	0.004	-0.005	0.010	0.010	0.011	0.008	0.008	0.015	-0.011	0.004	-0.012	-0.004
Man Cloth Spending	-0.029	-0.024	0.000	0.006	0.003	0.000	0.006	0.002	0.002	0.003	0.961	-0.013	-0.038	-0.023
Female Cloth Spending	-0.026	-0.023	-0.001	0.002	0.004	0.001	0.007	0.003	0.003	0.005	-0.037	-0.011	0.964	-0.020

Table 10. Canadian Household Spending on Private Health Insurance: A Post-VariMax Rotation Analysis (continued)

lg EP007	raj	Coef.	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19	Comp20	Comp21	Comp22	Comp23	Comp24	Comp25	Comp26
			-0.0177	0.0632	-0.0478	0.0143	-0.0027	-0.0523	0.0560	0.0273	-0.0339	-0.0039	0.0325	0.0344	0.0070
Prv_BC			-0.039	0.082	0.913	-0.003	-0.001	-0.007	0.016	0.003	-0.005	0.004	0.002	0.003	0.001
Prv_PEI			0.939	0.045	-0.029	0.001	-0.001	-0.007	-0.014	-0.009	0.003	-0.003	-0.004	-0.003	-0.002
Prv_NS			-0.081	0.077	-0.061	0.001	-0.001	-0.008	-0.010	-0.008	0.001	-0.002	-0.004	-0.001	-0.001
Prv_NB			-0.096	0.080	-0.058	0.001	-0.003	-0.011	-0.017	-0.013	0.003	-0.004	-0.006	-0.003	-0.003
Prv_QBC			-0.154	0.320	-0.342	0.001	0.002	-0.007	0.047	0.009	-0.022	0.004	0.003	0.013	0.004
Prv_MA			-0.052	-0.920	-0.078	-0.005	-0.003	0.002	0.006	0.003	-0.002	0.002	0.001	0.002	0.001
Prv_SA			-0.082	0.075	-0.057	-0.004	-0.002	0.001	-0.016	-0.002	0.005	0.000	-0.001	-0.002	0.000
Prv_AB			-0.044	0.131	-0.148	-0.002	0.002	0.007	0.035	0.014	-0.013	0.007	0.006	0.009	0.004
Prv_NEW			-0.087	0.072	-0.051	0.005	0.003	-0.009	-0.016	-0.007	0.002	-0.003	-0.004	-0.002	-0.001
1 Person Household			-0.014	0.015	-0.009	-0.291	-0.284	0.021	-0.051	-0.083	0.122	0.001	-0.012	-0.004	-0.046
Couple With Kid(S)			-0.003	0.011	-0.004	-0.150	-0.140	0.012	-0.016	-0.031	-0.025	-0.002	-0.004	-0.007	-0.005
Couple With Other			0.002	0.005	-0.003	-0.042	-0.037	0.009	-0.005	-0.001	-0.006	0.000	0.002	-0.001	0.001
Lone Parent			-0.002	0.003	-0.002	-0.060	0.940	0.003	-0.010	-0.021	0.016	0.000	-0.004	-0.002	-0.010
Other HH Type			0.300	0.005	-0.003	0.933	-0.062	0.011	-0.012	-0.010	0.019	0.001	0.001	0.000	-0.007
Own With Mortgage			-0.002	0.007	-0.009	0.016	0.016	0.030	-0.012	-0.036	0.008	0.029	-0.017	-0.024	-0.007
Miscellaneous Income Source			-0.002	-0.002	0.003	0.000	0.000	-0.004	0.000	0.002	-0.001	0.997	0.001	0.002	0.001
Self Employee			-0.003	-0.001	0.001	0.001	-0.003	-0.005	0.000	-0.014	-0.001	0.001	0.994	-0.004	-0.004
Investment Income			-0.002	-0.001	0.002	0.000	-0.002	-0.001	-0.002	-0.008	0.000	0.002	-0.004	0.997	-0.002
Government Payment			-0.008	-0.009	0.004	0.007	-0.026	-0.045	0.014	-0.094	-0.011	0.001	-0.038	-0.019	-0.023
Other Major Source Income			-0.008	-0.002	0.003	-0.006	-0.018	-0.014	0.000	0.960	0.000	0.002	-0.016	-0.010	-0.011
Suitable Place To Live			0.002	0.001	-0.003	0.014	0.012	-0.006	0.008	0.001	0.986	-0.001	-0.001	0.000	0.003
Have Internet			-0.001	-0.001	0.001	-0.005	-0.007	-0.002	-0.001	-0.009	0.003	0.001	-0.003	-0.002	0.997
Size of Household			0.002	-0.005	-0.001	0.079	0.076	0.004	-0.004	0.032	0.080	0.001	0.011	0.014	-0.006
Length of Tenure			0.012	0.023	-0.027	-0.040	-0.016	0.072	-0.024	0.010	0.034	0.033	0.007	-0.013	0.000
House Selling Value			-0.009	-0.024	0.037	-0.009	-0.014	-0.071	0.020	0.074	-0.025	-0.062	0.031	0.047	0.017
Higher Level of Edu			0.017	0.014	-0.021	-0.017	-0.004	0.026	0.004	0.029	0.007	0.006	0.013	0.006	0.006
Urban Size			-0.249	0.030	0.077	0.019	0.009	-0.026	-0.120	-0.041	0.031	-0.017	-0.019	-0.023	-0.006
Building Age			-0.006	-0.002	0.003	0.009	0.002	0.985	0.004	-0.012	-0.006	-0.004	-0.006	-0.001	-0.003
Income			-0.014	-0.026	0.007	0.062	-0.023	-0.119	0.048	-0.225	-0.041	-0.001	-0.087	-0.041	-0.049
Rec Car Insurance			-0.011	-0.003	0.011	-0.009	-0.007	0.004	0.987	-0.001	0.008	0.000	0.000	-0.002	-0.001
Man Cloth Spending			-0.003	0.002	0.001	-0.046	-0.041	0.010	-0.013	-0.005	0.019	0.001	0.001	0.000	-0.005
Female Cloth Spending			-0.004	0.002	0.002	-0.042	-0.038	0.010	-0.013	-0.007	0.019	0.001	0.001	-0.001	-0.005

Note: Table 10 presents the results of a VariMax rotation analysis for private health insurance spending. The table includes various variables and their corresponding coefficients across ten components (Cmp1 to Cmp10). Table 10's highlights key factors influencing private health insurance spending through components 1 to 10. Component 1 indicates that larger households, especially those with couples and children, tend to spend more on private health insurance, whereas single-person households spend less. Component 2 highlights that longer residence duration and higher property values are associated with increased spending, alongside a positive relationship with government payments. Component 3, however, shows a negative relationship with private health insurance spending, with households receiving government payments and lower incomes spending less on private health insurance.

Table 11. Canadian Household Data: An Analysis Post-VariMax Rotation for Car Insurance Holders

lg_EP007_raj Coef.	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12	Comp13
Variable	0.1057	0.0426	-0.1025	0.0214	-0.0022	-0.1183	0.0613	0.0470	0.0387	-0.0236	0.0582	0.0171	0.0122
Prv_BC	-0.006	-0.015	0.003	-0.107	0.001	-0.114	0.886	-0.004	0.066	0.064	0.071	-0.032	0.010
Prv_PEI	0.001	0.007	-0.004	-0.022	-0.003	-0.031	-0.033	0.003	0.065	0.065	0.062	0.017	-0.005
Prv_NS	-0.002	0.005	-0.003	-0.056	-0.002	-0.067	-0.068	0.002	0.092	0.091	-0.910	0.012	-0.004
Prv_NB	-0.001	0.009	-0.007	-0.055	-0.003	-0.067	-0.068	0.003	-0.896	0.102	0.100	0.019	-0.005
Prv_QBC	-0.006	-0.004	0.004	-0.152	0.019	0.837	-0.152	-0.004	0.084	0.080	0.091	-0.037	0.019
Prv_MA	-0.005	-0.009	0.002	-0.072	-0.001	-0.078	-0.080	-0.005	0.065	0.064	0.067	-0.014	0.002
Prv_SA	-0.016	-0.021	0.004	-0.373	-0.018	-0.381	-0.310	-0.006	0.352	0.340	0.331	0.000	-0.008
Prv_AB	-0.011	-0.028	0.015	0.805	0.014	-0.193	-0.178	-0.010	0.063	0.060	0.075	-0.067	0.024
Prv_NEW	0.000	0.006	-0.003	-0.050	-0.004	-0.061	-0.063	0.004	0.099	-0.902	0.096	0.017	-0.005
1 Person Household	-0.350	-0.131	-0.078	0.032	0.087	0.023	0.005	-0.077	0.010	0.001	0.007	-0.047	-0.202
Couple With Kid(S)	0.696	-0.048	-0.032	0.002	0.026	0.003	-0.001	-0.232	-0.002	-0.007	-0.001	-0.026	-0.075
Couple With Other	-0.025	-0.017	0.000	-0.004	0.007	-0.002	-0.003	0.934	-0.003	-0.004	-0.002	-0.015	-0.022
Lone Parent	-0.041	-0.019	-0.017	0.002	0.018	0.000	-0.002	-0.044	0.002	-0.001	0.001	-0.011	-0.044
Other HH Type	-0.045	-0.032	-0.007	0.001	0.017	0.002	-0.002	-0.050	-0.001	-0.003	-0.001	-0.021	-0.049
Own With Mortgage	0.021	-0.109	-0.021	0.006	0.850	0.027	-0.004	0.011	-0.004	-0.002	-0.004	-0.013	-0.001
Miscellaneous Income Source	0.000	0.010	0.000	0.002	0.011	0.000	0.003	0.000	0.002	0.001	0.001	0.005	0.001
Self Employee	0.003	0.008	-0.015	0.003	-0.007	0.001	0.000	0.003	0.004	0.002	0.002	0.011	0.002
Investment Income	0.001	-0.001	-0.008	0.005	-0.010	0.005	0.003	0.000	0.002	0.001	0.001	0.006	-0.002
Government Payment	0.027	0.095	0.824	0.012	-0.032	-0.009	-0.003	0.022	0.019	0.005	0.010	0.075	0.013
Other Major Source Income	0.000	0.009	-0.032	0.008	-0.016	0.005	0.003	-0.001	0.009	0.006	0.006	0.019	-0.006
Suitable Place To Live	0.010	0.015	-0.001	-0.007	0.000	-0.008	-0.005	-0.004	-0.002	-0.002	-0.001	0.006	0.019
Have Internet	0.000	0.004	-0.007	0.001	-0.001	0.000	0.001	0.002	0.000	-0.001	-0.001	0.004	0.000
Size of Household	0.613	-0.013	0.037	-0.008	0.021	-0.009	-0.010	0.234	0.005	0.009	0.006	-0.017	-0.007
Length of Tenure	-0.024	0.792	0.028	-0.019	-0.152	0.018	-0.023	-0.027	-0.024	-0.017	-0.016	-0.082	-0.044
House Selling Value	-0.027	0.514	0.042	0.003	0.484	-0.056	0.020	-0.009	0.030	0.024	0.024	0.060	0.023
Higher Level of Edu	-0.012	-0.035	0.022	-0.029	-0.001	-0.022	-0.024	-0.014	-0.016	-0.015	-0.010	0.958	-0.007
Urban Size	0.035	0.070	-0.032	0.400	-0.065	0.276	0.201	0.030	0.164	0.171	0.121	0.180	-0.058
Building Age	0.006	0.023	-0.011	0.003	0.009	-0.003	0.001	0.007	0.009	0.007	0.007	0.018	0.007
Income	0.085	0.239	-0.553	0.007	-0.058	-0.039	-0.020	0.069	0.033	0.006	0.016	0.142	0.052
Rec Car Insurance	-0.005	-0.006	0.001	0.021	-0.006	0.021	0.018	-0.003	0.010	0.011	0.007	0.009	-0.014
Man Cloth Spending	-0.019	-0.015	-0.001	0.007	0.006	0.008	0.005	-0.017	0.001	0.001	0.000	-0.007	-0.025
Female Cloth Spending	-0.023	-0.021	-0.003	0.010	0.003	0.012	0.007	-0.019	0.004	0.004	0.003	-0.007	0.968

Table 11. Canadian Household Data: An Analysis Post-VariMax Rotation for Car Insurance Holders (continued)

Ig EP007 raj Coef.	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19	Comp20	Comp21	Comp22	Comp23	Comp24	Comp25	Comp26
Variable	0.0599	0.0428	0.0116	-0.0239	0.0356	0.0078	-0.0227	-0.0592	0.0050	0.0057	-0.0085	-0.0167	0.0190
Prv_BC	-0.095	-0.003	-0.003	-0.045	0.008	0.025	0.002	0.004	-0.007	0.004	0.001	0.005	0.001
Prv_PEI	-0.044	0.002	0.000	0.945	-0.001	-0.013	-0.009	-0.008	0.003	-0.002	-0.003	-0.003	0.001
Prv_NS	-0.074	0.001	-0.002	-0.075	0.000	-0.009	-0.010	-0.008	0.002	-0.001	-0.003	-0.001	0.001
Prv_NB	-0.078	0.001	-0.002	-0.085	-0.001	-0.013	-0.013	-0.012	0.003	-0.002	-0.005	-0.003	0.001
Prv_QBC	-0.123	0.000	-0.002	-0.054	0.014	0.037	-0.004	0.009	-0.014	0.001	0.002	0.010	0.001
Prv_MA	0.926	-0.004	-0.004	-0.049	0.002	0.008	-0.001	0.001	-0.002	0.002	0.002	0.002	0.002
Prv_SA	-0.281	-0.005	-0.006	-0.228	0.000	-0.013	-0.013	-0.011	0.004	0.003	-0.002	-0.001	0.006
Prv_AB	-0.131	-0.004	-0.002	-0.034	0.016	0.052	0.008	0.020	-0.016	0.006	0.006	0.013	0.002
Prv_NEW	-0.074	0.003	0.001	-0.082	-0.001	-0.015	-0.011	-0.008	0.003	-0.002	-0.003	-0.002	0.002
1 Person Household	-0.003	-0.303	-0.297	-0.003	-0.163	-0.053	0.023	-0.069	0.143	0.001	-0.010	-0.007	-0.021
Couple With Kid(S)	-0.010	-0.173	-0.163	0.001	-0.068	-0.017	0.017	-0.030	-0.019	-0.001	-0.003	-0.006	0.002
Couple With Other	-0.004	-0.048	-0.043	0.003	-0.020	-0.004	0.009	-0.002	-0.007	0.000	0.003	0.000	0.003
Lone Parent	-0.004	-0.068	0.932	0.000	-0.038	-0.010	0.005	-0.018	0.021	-0.001	-0.003	-0.002	-0.004
Other HH Type	-0.004	0.926	-0.070	0.002	-0.042	-0.013	0.012	-0.010	0.024	0.001	0.002	0.000	-0.002
Own With Mortgage	-0.002	0.020	0.024	0.001	0.008	-0.012	0.027	-0.035	0.006	0.030	-0.015	-0.026	-0.001
Miscellaneous Income Source	0.002	0.001	-0.001	-0.002	0.000	0.000	-0.005	0.002	-0.002	0.996	0.001	0.002	0.000
Self Employee	0.000	0.002	-0.002	-0.003	0.002	0.001	-0.006	-0.014	-0.002	0.001	0.992	-0.005	-0.004
Investment Income	0.002	0.000	-0.002	-0.002	0.000	-0.003	-0.001	-0.009	0.001	0.002	-0.005	0.996	-0.002
Government Payment	0.003	0.013	-0.025	-0.010	0.013	0.016	-0.046	-0.102	-0.015	-0.001	-0.057	-0.029	-0.027
Other Major Source Income	0.001	-0.006	-0.014	-0.007	-0.003	-0.003	-0.009	0.969	0.001	0.002	-0.016	-0.010	-0.007
Suitable Place To Live	-0.002	0.017	0.015	0.002	0.015	0.010	-0.006	0.002	0.981	-0.002	-0.002	0.001	0.001
Have Internet	0.001	-0.001	-0.003	0.001	0.000	0.001	-0.003	-0.006	0.001	0.000	-0.004	-0.002	0.998
Size of Household	0.000	0.069	0.067	0.000	0.000	-0.003	0.003	0.029	0.095	0.001	0.013	0.013	-0.006
Length of Tenure	-0.015	-0.044	-0.019	0.018	-0.029	-0.019	0.066	0.000	0.037	0.038	0.011	-0.014	0.008
House Selling Value	0.007	-0.006	-0.022	-0.020	0.001	0.021	-0.071	0.068	-0.028	-0.067	0.026	0.050	0.003
Higher Level of Edu	-0.013	-0.018	-0.009	0.017	-0.007	0.010	0.020	0.020	0.006	0.006	0.012	0.007	0.005
Urban Size	0.060	0.024	0.019	-0.162	-0.031	-0.148	-0.036	-0.055	0.038	-0.011	-0.016	-0.032	0.004
Building Age	0.000	0.009	0.004	-0.008	0.006	0.002	0.988	-0.008	-0.006	-0.005	-0.006	-0.001	-0.003
Income	0.001	0.062	-0.014	-0.013	0.047	0.046	-0.092	-0.181	-0.041	-0.004	-0.104	-0.048	-0.050
Rec Car Insurance	0.007	-0.009	-0.007	-0.011	-0.010	0.983	0.002	-0.004	0.010	0.000	0.001	-0.003	0.001
Man Cloth Spending	0.002	-0.032	-0.029	-0.001	0.980	-0.011	0.006	-0.004	0.015	0.000	0.002	0.000	0.000
Female Cloth Spending	0.002	-0.039	-0.036	-0.004	-0.026	-0.015	0.008	-0.007	0.020	0.001	0.001	-0.002	-0.001

Note: Table 11. presents a Post-VariMax rotation analysis of car insurance spending among Canadian households, focusing on the first three components. Component 1 reveals that households with couples and children (loading: 0.70) and larger households (loading: 0.61) are associated with higher car insurance spending, while single-person households (loading: -0.35) tend to spend less. Component 2 indicates that households with longer residence durations (loading: 0.79) and higher property values (loading: 0.51) tend to spend more on car insurance. Component 3 shows a negative relationship with car insurance spending, where lower household incomes (loading: -0.55) and those receiving government payments (loading: 0.82) are associated with lower car insurance expenditures. These findings suggest that household composition, financial stability, and income levels significantly influence car insurance spending patterns among Canadian households.