THE INFLUENCE OF UTILITY FUNCTIONS ON INSURANCE CHOICES

by

Mingzhu Wang

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
April 2018

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols Used

pm	Profit margin
APV	actuarial present value
ARA	Absolute risk aversion
GAM	Generalized additived models
PCA	Principal component analysis
P	Premium price
SVD	Singular value decomposition

Abstract

As society advances, people's quality of life is improving. More and more people are not only paying attention to the physical quality of life but also focusing on the wealth management such as buying insurance to support their life quality. Because of this, insurance companies are providing a growing number of policies to satisfy the public need. Utility is a measurement of people's welfare. We aim to estimate a utility function model based on the people's preferences. However, the following reasons make the utility function hard to estimate: First of all, for existing models, it is difficult to apply these models to describe multifarious people's preferences. Secondly, it is difficult to obtain the exact utility model from datasets. The primary issue is how to establish a good predictive model for all different insurance needs. In this thesis, we are trying to develop a framework for selecting suitable insurance choices. We compute the amount of insurance to purchase for a range of randomly generated utility functions and a range of situations. We then apply data mining approaches, such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) to determine which aspects of the utility function are most influential on insurance choice, and use these aspects to reduce the dimension of the predictive model. Finally, we test the reduced dimension of predictive models on the simulated datasets. Our results demonstrate that the some reduced predictive models can get high predictive accuracy under a range of conditions.

Acknowledgements

With my most sincere gratitude and heartfelt, I would like to thank all of the people who have given me support and suggestion to help me complete this thesis.

I would like to thank Prof. Toby Kenney who is my thesis and project supervisor. From the choice of subject to the final completion of the project, Prof. Toby Kenney has always given me careful guidance, not only set me a lofty academic goal, but also made me understand many truths of real life. My sincere thanks to Prof. Hong Gu and Prof. Bruce Smith for reading and giving constructive comments on this thesis. I would say, I will always remember these times and amazing experience.

In life and study, I also got the concern and help from friends and classmates. Last but not the least, I must express my gratitude to my parents for supporting me spiritually throughout this program. Without their support it would not be possible to conduct this research.

Chapter 1

Introduction

1.1 Insurance Background

1.1.1 Insurance History

Since the begining of time, human beings have always been threatened by natural disaster, illness and accidents. In order to protect themselves from disasters, the idea of insurance against the various natural disasters and accidents and illness emerged [5].

Among the various types of insurance, the earliest was marine insurance. Because of the development of marine insurance, the entire insurance industry had been improved. At that time, there was a common view among merchants that the losses caused by the abandonment of goods for the safety of cargo should be shared among the beneficiaries. This view is being recognized as the seed of marine insurance. In the late 15th century, the first life insurance policies were on African slaves being transported as goods to the Americas [19].

With the continuous development of the world economy and the improvement of people's lives, an increasing number of people is starting to pay attention to the protection of personal growth and property value. Hence, insurance has been created to protect people from financial loss. Modern insurance industries have emerged and have been developing rapidly. Currently, there are thousands of insurance companys, and hundreds of insurance products available. The main products are life insurance, health insurance, property insurance, and so on [7]. In this thesis, I will focus on the life insurance product.

1.1.2 Life Insurance

Life insurance is a contract where an insurance company offers to provide a monetary benefit to the person who bought the insurance product in the event of the death of the insured. That person is often called the policyholder. When the policyholder purchases an insurance product, he or she needs to pay the premium to the company until the contract ends. If the policyholder dies during the contract period, the contract would stop immediately, and the company should provide a monetary benefit to the policyholder's estate.

1.2 Utility and Utility Function Analysis

1.2.1 Utility

The utility is a number which measures the satisfaction an individual derives from something. In this thesis, we will be considering utility as a function of wealth. This is a commonly used model in economics to explain financial choices. The utility function is not directly observable, and some research [3] has suggested that it provides a poor model of how people actually make choices. While, utility has found substantial application in the economics literature, e.g.welfare [15] for explaining the decisions made by individuals, it is less frequently used in practice to help with these decisions. The reason is the difficulty in measuring the utility function.

1.2.2 Utility Function

Utility typically depends on a number of variables, with wealth being the most important one. We therefore describe the utility as a function of wealth. Since wealth is a random variable, the utility function is also a random variable, and maximizing the expected value of utility is often taken as an individual's objective.

1.2.3 The St Petersburg Paradox

In the context of actuarial science, money is focused on the utility to measure the agent's satisfaction. The first suggestion of utility function was from Daniel Bernoulli in 1738, in his solution to the St Petersburg Paradox [11].

In the 18th century St Petersburg, Bernoulli considered a game, which involved tossing a fair coin repeatedly and paying some money to the casino, he would stop when the first "tail" appeared. The payoff for the game is 2^{N-1}, where N is the number of times the coin was tossed. [8] [16]. The expected payoff of this game is infinite.

$$E[X] = \sum_{n=1}^{\infty} 2^{n-1} \frac{1}{2^n} = \sum_{n=1}^{\infty} \frac{1}{2} \tag{1.1}$$

That means no amount of money is too large to pay for one game.

Bernoulli suggested that the utility of an individual was a logarithmic function of the amount of money, such as

$$u(w) = k \log(w) + c \tag{1.2}$$

where k is a parameter, and w represents the wealth or the amount of money, c is a constant value. Under this hypothesis, an individual would choose to pay to play the game if the cost c increases the expected utility. For example, if the individual's starting wealth is w_0, and the cost of the game is p, the expected utility if the individual does not play is $k \log(w_0) + c$, and if the individual plays, it is $\sum_{n=1}^{\infty} \frac{1}{2^n} \left(k \log(w_0 - p + 2^{n-1}) + c\right)$, so the individual would be willing to pay p provided

$$\sum_{n=1}^{\infty} \frac{\log(w_0 - p + 2^{n-1})}{2^n} > \log(w_0) \tag{1.3}$$

The utility form is obviously increasing, $u'(w) > 0$, as it is easy to give away the additional wealth. The key feature of the logarithm function chosen by Bernouill is that it is concave, $u''(w) < 0$, which is what economists refer to as *decreasing marginal utility.*

1.3 Basic Data Mining Methods

If we are given an individual's utility function, it is straight-forward to calculate the value of an insurance choice (and thus, to decide which choice is more appropriate. More details are given in Chapter 2.2.1). However, in practice, we do not know the individual's utility function and the utility function is very hard to determine. The aim of this thesis is to simplify the problem of estimating the utility function to a small number of key features, that best predict an individual's insurance choices. We will begin with a large number of candidate variables, and use two data-mining techniques to identify the key features. Both are intended to remedy the issues caused by high-dimensional data. The first is ridge regression [9]. This is a method for estimating regression parameters in a high-dimensional case where noise is much larger than signal. It is a penalised least-squares approach with penalty term $\lambda\|\beta\|^2$. The tuning parameter λ is a model complexity determinant: as the value of λ increases, the complexity reduces. We use cross-validation to choose the optimal λ value.

There are a number of other methods that can estimate parameters in this high-dimension case, of which the most popular one is Lasso [18]. In Lasso, we have the penalty term $\lambda\|\beta\|_1$ and λ works similarly to that of ridge regression. But Lasso will set the less-important predictors to zero and the model becomes more sparse [13]. In this thesis, we want the coefficients to vary smoothly not sparsely, thus ridge regression seems more appropriate.

The other data mining technique we use is Singular Value Decomposition (SVD). This is a matrix decomposition $X = UDV^T$ where U and V are orthogonal matrices, D is a diagonal matrix with non-negative values on the diagonal. The idea is that by restructuring to the first few singular vectors, we are able to get a good low-rank approximation of X and thus reduce the dimension [4]. In our case, X is the matrix of coefficients estimated by ridge regression.

1.4 Structure of Thesis

The goal of this thesis is to find what kind of measurement of the utility function is needed to find the correct insurance choices for an individual. The remainder of this thesis is divided into 4 chapters. Utility function analysis on common parametric utility functions and information on the measurement of risk aversion are described in Chapter 2. Chapter 3 conducts a large-scale simulation study to identify the key features. We devote Chapter 4 to interpretation of the key features identified. The thesis conclusion is in Chapter 5.

Chapter 2

Utility Function Selection

2.1 Utility Theory

Utility theory is to describe an individual's behaviour under uncertain conditions, and the critical content is the utility function $u(w)$.

In general, utility refers to a satisfiaction level when consumers are buying products. Decision-makers have different feedback from a specific event, some of them may be interested in the expected return, some may care more about avoiding losses than about the chance to gain. Mathematically, we represent this via a utility function $u(w)$, and assume each individual seeks to maximize the expectation of their utility.

Consider three investment policies:

Policy A		Policy B		Policy C	
x	$p(x)$	x	$p(x)$	x	$p(x)$
6	1	-5	1/4	-10	1/5
		0	2/4	0	1/5
		50	1/4	10	1/5
				20	1/5
				30	1/5

Table 2.1: Example of Three Investment Policies

We can calculate the expected return over these three policies.

$$E_A(x) = 6(1) = 6$$

$$E_B(x) = -5(1/4) + 0(2/4) + 50(1/4) = 9$$

$$E_C(x) = -10(1/5) + 0(1/5) + 10(1/5) + 20(1/5) + 30(1/5) = 10$$

The maximum expected return is from policy C. However, some investors may prefer policy A, because of the certain gain of 6. Some may prefer other policies, that will depend on the risk attitude of individuals. There are three types of risk attitudes, which have the following characteristics.

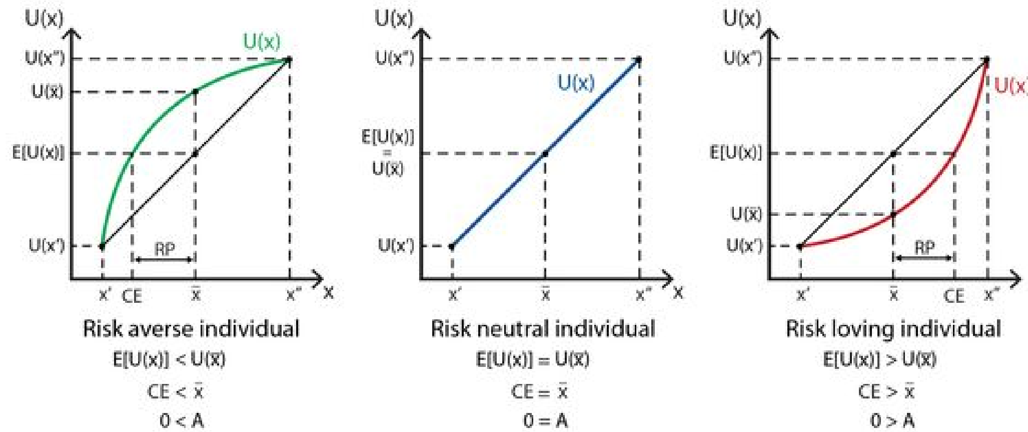

Figure 2.1: Risk preferences. The x-axis represents the income, and the y-axis is utility. [14]

- Risk-averse: The expected utility is less than or equal to the utility of expected value, $E(u(w)) \leq u(E(w))$ [12]. By Jensen's inequality, this holds if $u(w)$ is a concave function. We refer to u' as monotone decreasing marginal utility.

- Risk-neutral: The expected utility is equal to utility of expected value, $E(u(w)) = u(E(w))$, which is the case when $u' > 0, u'' = 0$. Here, u' is constant marginal utility. And $u(w)$ is a monotone increasing linear function.

- Risk-seeking: The expected utility is greater than or equal to the utility of expected value, $E(u(w)) \geq u(E(w))$. This happens when $u' > 0, u'' > 0$. So, u' is monotone increasing marginal utility. $u(w)$ is a monotone increasing convex function.

Figure 2.1 shows some graphical intuition about three types of risk preferences. For example, consider a situation where the individual has a choice between a certain gain of $\bar{x}$, and a gamble which pays x' or x'' with equal probability (where $\bar{x} = \frac{x'+x''}{2}$).

In the plots, the black line represents the expected utility of a gamble. For the risk averse individual, u is concave, the level of expected utility for the gamble is lower than the utility of $\bar{x}$ with certainly, so the certain $\bar{x}$ is prefered. For the risk neutral individual, u is linear, the level of expected utility for the gamble is the same as for $\bar{x}$ with certainly, so this individual is indifferent between these choices. In the risk loving individual, u is convex, so the gamble is preferred.

For the first plot of Figure 2.1, at wealth value $\bar{x}$, the utility of this certain value is higher than the utility of the lottery, $U(\bar{x}) > E[U(x)]$, the difference between $U(\bar{x})$ and $E[U(x)]$ is how risk-averse the person is: the difference between the expected utility of the gamble and the utility of getting that same amount of money for certain. To sum up, if there is some way to remove the risks from the gamble, this person would be happier. On the other hand, adding risks makes this person less happy when they are risk-averse. This explains why people use insurance companies. Insurance companies remove some risk from life; for instance, if your house burns down, you're going to suffer a big loss, so you're not certain. Buying insurance will reduce your income with certainty, but then there is a chance that if you get the bad outcome, someone will step in and fix your house. The risk-averse individual would be willing to pay some money in order to remove this risk. However, how much would they be willing to pay? From Figure 2.1 first plot, CE is certainty equality, which is the amount of money that has the same utility as the expected utility of the gamble, so this person values the gamble at CE if they could remove the uncertainty. The horizontal black dashed line at $E[U(x)]$, and the distance between the green line and black line is called risk premium(RP), which is the price that this person would pay to reduce risks. Insurance companies are insuring a lot of different risks; some of them will work out in the favor of insurance company, and some of those won't work out in the favor of the insurance company. If they have enough people insured, then the relative risk is reduced, and the insurer can act as if they are risk netural. Therefore, in long-run, the insurance policies can achieve a win-win situation. [1]

In general, the decision-makers are risk averse, and their utility curves are increasing and concave. In this thesis, we suppose the agents are all risk averse. We need to consider inflation. In this case, we set a parameter called time decay, which is a factor that weights future utilities less than current utility. We use exponential time decay, i.e. we multiply utility by the factor $e^{-\tau t}$, where t is time and τ is a parameter that varies between individuals.

The most widespread measurement of risk aversion is *absolute risk aversion*(ARA),

$$R_a(w) = -\frac{u''(w)}{u'(w)} = -\frac{\frac{d^2u}{dw^2}}{\frac{du}{dw}} \tag{2.1}$$

where $u'(w)$ and $u''(w)$ are the first and second derivatives with respect to w of utility function $u(w)$, where w represents the individual's wealth.

2.2 Parametric Utility Models and Estimation

It is difficult to get what the utility function actually is of each agent, because utility is not directly quantifiable. One common practice is to set a parametric form of the utility function.

Common choices of this parametric form are:

- Logarithmic utility: $u(w) = k\log(w) + c$, c and k are parameters. [6]
- Exponential utility: $u(w) = -ke^{-\alpha w},\ \alpha > 0$ and k are parameters. [10]
- Power utility: $u(w) = \frac{w^{\alpha}-1}{\gamma},\ \gamma < 1$, α is a non-negative parameter. [2]

We will now look at some examples of how these utility functions are applied to making insurance choices. To describe the scenarios in our examples, we will need the following concepts.

Definition 1. *Premium is a payment to an insurance company from the contract with a fixed period, like monthly.*

Definition 2. *Mortality is the instantaneous dealth rate at a certain age, measured on a specific time basis, such as annually.*

Definition 3. *Death benefit is a payment to the beneficiary upon the death of the policyholder.*

In all our scenarios, a potential policyholder will be deciding how much life insurance to buy based on their utility function.

We demonstrate the effects utility function can have on insurance choices in a simple scenario. We suppose an individual's current wealth is 1, and wealth is increasing continuously at a rate of 0.01 while the individual is alive, and will decrease at rate 0.1 after the individual dies. We suppose the individual buys p units of a life insurance policy with a continuous premium rate 0.01 and dealth benefit 0.8. At time t, the policyholder's wealth will be

$$w = \begin{cases} 1 + 0.01t - 0.01pt & t < s \\ 1 + 0.01s - 0.01ps + 0.8p - 0.1(t-s) & t > s \end{cases} \tag{2.2}$$

where s is the future death time of the policyholder, t is the current time, and p is the amount of insurance purchased. We will assume constant mortality of 0.01, so s is exponentially distributed with mean 100.

2.2.1 Calculating Insurance Choice for Each Wealth, Policy and Utility

Given a model of wealth development, a model of mortality, an insurance policy and a utility function, we want to compute how much insurance the individual should purchase. We let the objective function be a weighted average of the utility at all times with time decay 0.058. That is, if the wealth at time t is $w(t)$. The utility is $\int_0^t e^{-0.058t} u(w(t))dt$

Now we consider using the expected value of utility to find the optimal amount of insurance to buy. The expected utility is given by

$$\mathbf{E} = \int_0^{\infty} \int_0^{T} u(w(P, t, s, y), t, \tau)\ m(s)\ dtds \tag{2.3}$$

where s is the future death time of the policyholder, t is current time, P is the premium, y represents the amount of insurance the policyholder bought, τ is the factor for time decay, T is the term of the contract. u is the utiliy function, w is wealth, mortality usually refers to force of mortality, but here m is the probability density of time until death.

We analyse three parametric utility functions, denoted u_1, u_2, u_3, given by $u_1 = log(\lambda w), u_2 = -e^{-\lambda w}, u_3 = (w+1)^{\frac{\frac{1}{\lambda}-1}{\rho}}$ with $\lambda = 4$ and $\rho = 0.9$.

Figure 2.2 shows the expected utility for the three parametric utility functions u_1, u_2, and u_3. From the left side utility plots, the three utility function curves show very different shapes of expected utility, and most importantly, lead to very different insurance choices. The choice of parametric model has a large effect on insurance choice. The right hand plots show the effect of parameter values on the insurance choices. Even for a parametric utility function, it can be difficult to estimate the parameters. In this example, we imagine that we are able to measure the utility at the points 0.5, 1.0, 1.5, but that our measurements are subject to normal error with mean 0 and variance 1 at each of these points. We then use nonlinear least squares to estimate the parameter λ and calculate the expected utility under this estimated utility function. We see that estimating parameters has a small effect compared to

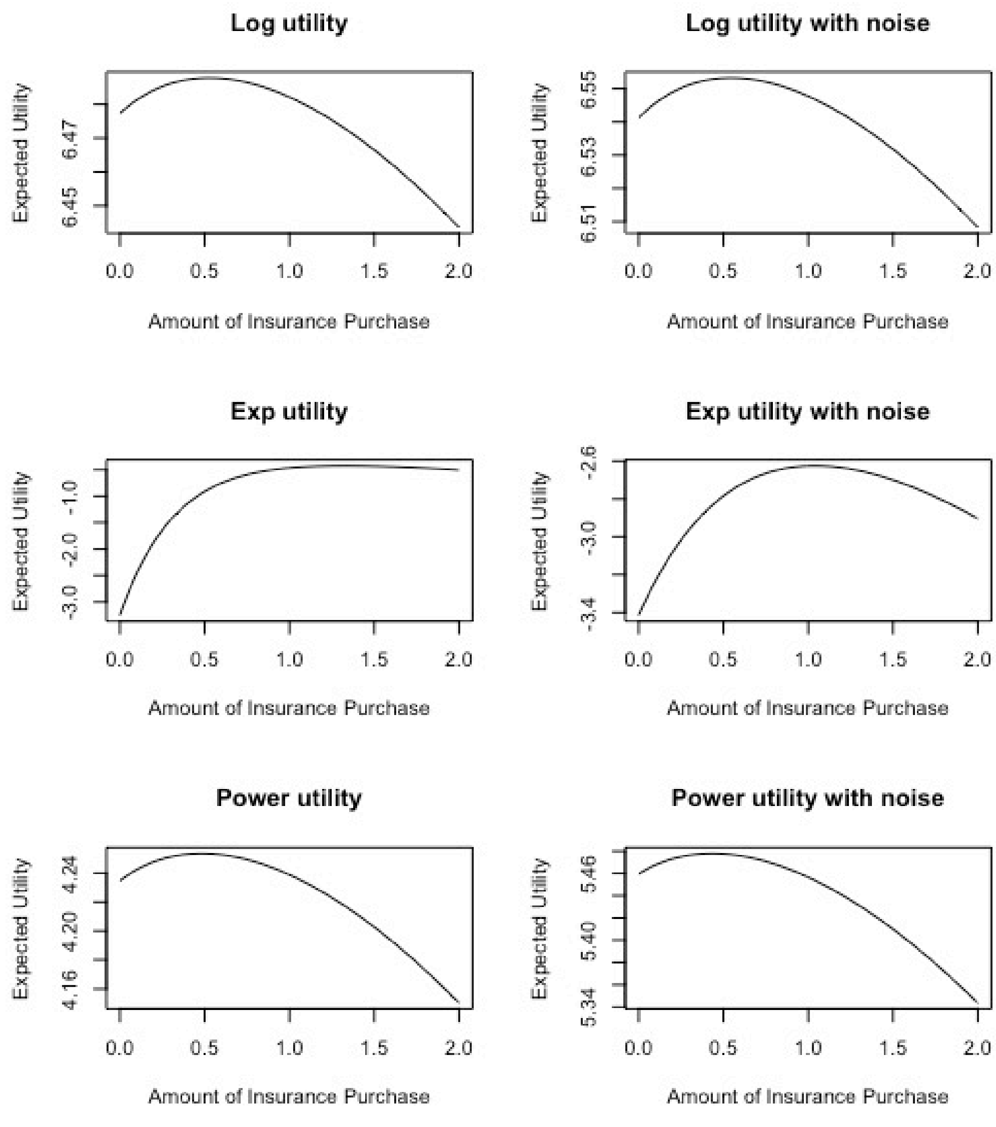

Figure 2.2: Expected value plot of parametric utility function: The left side plots are the exact parametric utlity functions and the right side plots are estimating parametric utility functions.

the difference between different parametric forms.

Chapter 3

Important Features of Utility Functions

In Chapter 2, we saw that the choice of parametric model has a large effect on the insurance purchase amount. This suggests that we cannot find the right insurance choice for an individual based on a 1-dimensional measurement of their utility. Given the difficulty of measuring an individual's utility function, it would still be desirable to find a low-dimensional description of the utility function which is able to predict an individual's insurance choices. Finding such a description is the aim of this chapter.

3.1 Overview of Simulation Design

We start by simulating the data from 840 scenarios, by combining 3 wealth development schemes, 4 loadings and 70 mortalities. We also generate 1000 random utility functions. For each combination of a scenario and a utility function, we calculate the optimal amount of insurance to purchase. This results in a 1000*840 matrix Y of insurance choices. For each utility function, We compute ARA at 699 values of w, to obtain a 699-dimensional description of the utility. This results in a 1000×699 predictor matrix X. We then fit regression models $Y = XB + E$ to predict insurance choice for each scenario, and use dimension reduction to reduce the coefficients from the regression model to a low-dimensional space. That is, we perform dimension reduction on the 699×840 matrix B, to get a low-rank approximation. For every utility function, its projection to this low-dimensional space should lead to the same insurance choice as the original utility function.

3.2 Non-Parametric Utility functions using Splines

Splines are a popular choice for non-parametric functions. A spline is a piecewise polynomial curve of degree $n-1$ with the pieces chosen so that the whole function is $n-2$ times continuously differentable. The spline regression model is a special case of a basis function approach: we can write a spline

$$f(x) \equiv \beta_0 + \beta_1\ b_1(x) + \beta_2\ b_2(x) + \beta_3\ b_3(x) + \cdots + \beta_k\ b_k(x) \tag{3.1}$$

where $b_1(\cdot), b_2(\cdot), \ldots, b_k(\cdot)$ are the fixed basis functions.

In this thesis, we use the B-spline basis for spline functions. The non-decreasing sequence of cutpoints which defines different regions are called knots, and denoted $s_1 \leq s_2 \leq \cdots \leq s_k$. We extend the sequence by adding additional knots $t_1 = t_2 = \cdots = t_n = s_1, t_{n+1} = s_2, \ldots, t_{n+k-1} = s_k, t_{n+k} = \cdots = t_{2n+k-2} = t_{n+k-1}$. We denote the ith B-spline basis function $B_{i,n}(x)$, and we can express the spline function as a linear combination of B-splines:

$$S_{n,t}(x) = \sum_i \beta_i\ B_{i,n}(x) \tag{3.2}$$

where S is the spline function we are describing. The B-spline basis is calculated based on the Cox-de Boor recursion formula:

$$B_{i,1}(x) = \begin{cases} 1 & if\ t_i \leq x < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

$$B_{i,n}(x) = \frac{x - t_i}{t_{i+n-1} - t_i} B_{i,n-1}(x) + \frac{t_{i+1} - x}{t_{i+n} - t_{i+1}} B_{i+1,n-1}(x) \tag{3.4}$$

In our case, we set n = 3, for quadratic splines and use equally spaced knots, with the distance d between them, so the B-spline basis is given by the formula

$$B_{1,3}(x) = \begin{cases} 1 + \frac{2(x-t_3)}{d} & x < t_3 \\ -\frac{(x-t_4)^2}{d^2} & t_3 \leq x < t_4 \\ 0 & x \geq t_4 \end{cases} \tag{3.5}$$

$$B_{2,3}(x) = \begin{cases} 2 & x < t_3 \\ 2 - \frac{(x-t_3)^2}{d^2} & t_3 \leq x < t_4 \\ \frac{(x-t_5)^2}{d^2} & t_4 \leq x < t_5 \\ 0 & x \geq t_5 \end{cases} \tag{3.6}$$

$$B_{i,3}(x) = \begin{cases} \frac{(x-t_i)^2}{d^2} & t_i \le x < t_{i+1} \\ \frac{3}{2} - \frac{2(x-\frac{t_{i+1}+t_{i+2}}{2})^2}{d^2} & t_{i+1} \le x < t_{i+2} \\ \frac{(t_{i+3}-x)^2}{d^2} & t_{i+2} \le x < t_{i+3} \\ 0 & \text{otherwise} \end{cases} \quad for\ 3 \le i < k-1 \tag{3.7}$$

and similar definitions for $B_{k,3}$ and $B_{k+1,3}$. We let the range of wealth be between 0 and 4, and we want to generate a random spline function which is increasing and concave on this interval. The number of internal knots k is simulated following a Poisson distribution with parameter $\lambda = 5$. In each simulation, the total number of knots is $k + 2$. The knots are evenly spaced in the interval [0, 4]. For example, if $k = 3$, the knots are at 0, 1, 2, 3 and 4. There are some constraints that can control the shape of the B-spline function. We want the random spline function to be increasing and concave. We assume the knots are equi-spaced at a distance d. The derivatives on the B-spline basis by the example formula:

$$B'_{1,3}(x) = \begin{cases} \frac{2}{d} & x < t_3 \\ -\frac{2(x-t_4)}{d^2} & t_3 \le x < t_4 \\ 0 & x \ge t_4 \end{cases} \tag{3.8}$$

$$B''_{1,3}(x) = \begin{cases} -\frac{2}{d^2} & t_3 \le x < t_4 \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

$$B'_{2,3}(x) = \begin{cases} 0 & x < t_3 \\ -\frac{2(x-t_3)}{d^2} & t_3 \le x < t_4 \\ \frac{2(x-t_5)}{d^2} & t_4 \le x < t_5 \\ 0 & x \ge t_5 \end{cases} \tag{3.10}$$

$$B''_{2,3}(x) = \begin{cases} -\frac{2}{d^2} & t_3 \le x < t_4 \\ \frac{2}{d^2} & t_4 \le x < t_5 \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

$$B'_{i,3}(x) = \begin{cases} \frac{2(x-t_i)}{d^2} & t_i \le x < t_{i+1} \\ -\frac{4\left(x-(\frac{t_{i+1}+t_{i+2}}{2})\right)}{d^2} & t_{i+1} \le x < t_{i+2} \\ \frac{2(x-t_{i+3})}{d^2} & t_{i+2} \le x < t_{i+3} \\ 0 & \text{otherwise} \end{cases} \quad for\ 3 \le i < k-1 \tag{3.12}$$

$$B''_{i,3}(x) = \begin{cases} \frac{2}{d^2} & t_i \le x < t_{i+1} \\ -\frac{4}{d^2} & t_{i+1} \le x < t_{i+2} \\ \frac{2}{d^2} & t_{i+2} \le x < t_{i+3} \\ 0 & \text{otherwise} \end{cases} \quad for\ 3 \le i < k-1 \tag{3.13}$$

The derivatives on the B-spline function are:

$$S'(t_i) = \frac{2}{d}\left(\beta_{i-1}(t_{i+1} - t_i) - \beta_{i-2}(t_{i+1} - t_i)\right) = 2(\beta_{i-1} - \beta_{i-2}) \tag{3.14}$$

$$S'(x) = \frac{1}{d}\sum_{i=2}(\beta_i - \beta_{i-1})\ B_{i,n-1}(x) \tag{3.15}$$

$$S''(x) = \frac{2}{d^2}\sum_{i=3}(\beta_i - 2\beta_{i-1} + \beta_{i-2})\ B_{i,n-2}(x) \tag{3.16}$$

When $S'(x) > 0$ and $S''(x) < 0$, the shape of B-spline is increasing and concave. A sufficient condition for $S'(x) > 0$ is $\beta_i > \beta_{i-1}$ for all i; Meanwhile a sufficient condition for $S''(x) < 0$ is $\beta_i - 2\beta_{i-1} + \beta_{i-2} < 0$, if we let $d_i = \beta_i - \beta_{i-1}$, then the conditions are $d_i > 0$ and $d_i < d_{i-1}$. We simulate $k+2$ values of d_i with these properties by simulating $d_0 = e^z$, where z $\sim$ N(0,1), $d_{i+1} \sim u(0, d_i)$. Finally, we set β_0 so that we have $\sum \beta_i = 1$

Figure 3.1 shows 10 randomly generated non-parametric utility functions. We want to compare these random utility functions over a large range of insurance policies and life situations.

3.3 Insurance Policies

We will use a simple term life insurance policy with a term of 20 years. There are three variables that might affect the policy: death benefit, policy premium and mortality. In the continuous case of term insurance, if the policyholder dies within an m-year term, the death benefit payment should be immediately paid to the insured. The present value of a \$1 of benefit at policyholder age x of 20-year term insurance is denoted $\tilde{Z}$,

$$\tilde{Z} = \begin{cases} v^{T_x} = e^{-\delta T_x} & if\ T_x \le 20 \\ 0 & if\ T_x > 20 \end{cases} \tag{3.17}$$

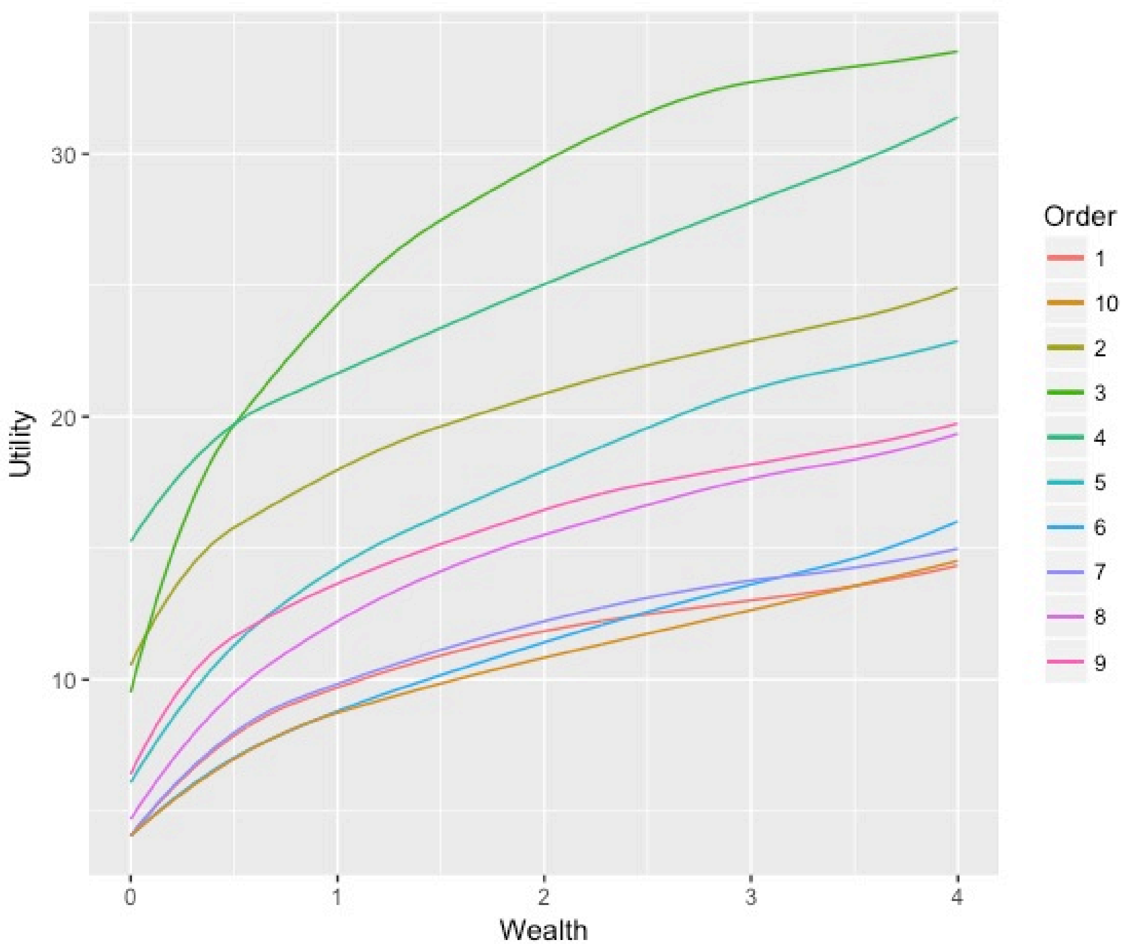

Figure 3.1: Ten Randomly Generated Non-Parametric Utility functions

where v is annual discount factor with $v = \dfrac{1}{1+i}$, i is the interest rate, T_x is the number of future years completed before the policyholder death, δ is the force of interest given by $\delta = \log(1+i)$.

The actuarial present value(APV) is given by

$$\bar{A}^{1}_{x:\overline{20}|} = \int_0^{20} e^{-\delta t}{}_tp_x\mu_{x+t}dt \tag{3.18}$$

where t is time, ${}_tp_x$ is the probability of a policyholder surviving from age x to age $x+t$, μ_{x+t} is the force of mortality at age $x+t$. The mortality is from the Society of Actuaries (SOA) life table and we use $i = 0.06$. [17]

In the term insurance, we consider the premium as a series of payments made by an individual at annual intervals, called an annuity. We use a discrete annuity to approximate a continuous annuity, which is simply an exchange of a lump sum of cash for a wealth development, the expected present value (EPV) of an annuity with payments \$1 can be determined as

$$\ddot{a}_{x:\overline{20}|} = \sum_{t=0}^{20-1} v^t{}_tp_x \tag{3.19}$$

In practice, the insurance company will add a percentage loading to the premium to cover profit and expenses. So, the premium P can be calculated as

$$P = \frac{\bar{A}^{1}_{x:\overline{20}|}}{(1-\zeta)\ \ddot{a}_{x:\overline{20}|}} \tag{3.20}$$

where ζ is the percentage loading.

We also apply a range of typical models for wealth development. Typically the policyholder will earn a salary while alive, and after the death of the policyholder, the policyholder's family will spend the wealth. This is the reason for buying life insurance.

Let s be the time at which the life dies, t be current time, and we set $0 < t < 20$, premium P is the price of one unit of the insurance policy and y is the amount of insurance bought. If the rate of salary minus expenditure while alive is $a(t)$ and the rate of expenditure after death is $e(t)$, then the wealth at time t for an individual who dies at time s is

$$w = \begin{cases} 1 + \int_0^t a(r)dr - Pyt & t < s \\ 1 + \int_0^s a(r)dr - Pys - \int_s^t e(r)dr & t > s \end{cases} \tag{3.21}$$

If $a(t)$ and $e(t)$ are constant, this gives a linear wealth.

We study 3 scenarios of wealth function. The first two scenarios are linear, where the salary and spending are at a constant rate. For a wealth scenario 3, we set expenditure and salary growing at an exponential rate. This can reflect realistic situations, such as inflation and career advancement.

$$w_1 = \begin{cases} 1 + 0.01t - Pyt & t < s \\ 1 + 0.01s - Pys + y - 0.2(t - s) & t > s \end{cases} \tag{3.22}$$

$$w_2 = \begin{cases} 1 + 0.05t - Pyt & t < s \\ 1 + 0.05s - Pys + y - 0.1(t - s) & t > s \end{cases} \tag{3.23}$$

$$w_3 = \begin{cases} 1 + 0.5e^{0.05t} - Pyt & t < s \\ 1 + 0.5e^{0.05s} - Pys + y - 0.5e^{0.05(t-s)} & t > s \end{cases} \tag{3.24}$$

Compared to scenario 1, scenario 2 represents on the one hand a larger salary, making insurance more affordable, and on the other hand, lower expenditure after death, making insurance less necessary. We use mortality from the SOA lifetable. To study the effect of different mortalities, we use different policyholder ages ranging from 21 to 90. We simulate four different values of loading: 0.15, 0.2, 0.25, and 0.3.

3.4 Method for Describing the Utility Functions

We want to summarize the utility function in terms of a collection of variables. We could use the values $u(w)$ for various w. However the notion of risk aversion is generally viewed as important in judging the effect of utility functions, so it is more natural to use this quantity

$$r(w) = -\frac{u''(w)}{u'(w)}$$

This has the advantage that it is invariant under translation and scaling of the utility function. We can compute the utility function up to translation and scaling from the risk aversion

$$r(w) = \frac{d}{dw} log(u'(w))$$

$$log(u'(w)) = \int r(w) dw$$

$$u(w) = \int e^{\int r(w)dw} dw$$

we evaluate the risk aversion using numerical differentiation at 699 equally-spaced values of w between 0 and 4. The predictor matrix X is obtained by concatenating the row vectors of ARA values for each utility function.

3.4.1 Ridge Regression Estimation

Suppose we have generated 1000 utility functions. Then in each of our 840 scenarios, we calculate the insurance choice in that scenario, hence we can set this table as a true choice dataset. According to the simulated insurance choice dataset, we would like to construct a regression model to predict the insurance choice. We can then look for a reduced-dimension model for this prediction. We could use ordinary least squares (OLS) for estimating the unknown parameters in a linear regression model to predict the insurance choice from the risk matrix. The predictor matrix X is obtained by concatenating the row vectors of ARA values for each utility function. We want to use a regression model to predict the insurance choice from the risk matrix. We fit one regression model for each scenario, obtaining one column of the matrix B at a time. But in this case, it is hard to evaluate a good estimated result from OLS because of the large number of variables. Penalised regression is a method for improving estimates in high-dimensional cases, by reducing the variance of the estimator. There are several common choices of penalty term. The most popular are Lasso and ridge regression. We choose ridge regression because we expect the coefficients to vary smoothly rather than being sparse.

We want to use a regression model to predict the insurance choice from the risk matrix. We would use y is a single column of the matrix Y from Section 3.1, β is the corresponding column of the matrix B. Suppose there are m observations $\{y_i, x_i\}_{i=1}^m$. For each observation i, there is a response y_i and predictors x_{ij} for $j = 1, ..., p$. The linear regression model is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{ip} + \epsilon_i,\ i = 1, ...,\ 840,\ j = 1, ...,\ 1000 \tag{3.25}$$

This model can be written in matrix notation as $Y = X\beta + \beta_0 + \epsilon$, where vector Y is in length 1000, X is an 1000×699 matrix, β is a vector of length 699.

In the OLS method, we usually minimize the residual sum of squares (RSS) to get the estimator $\widehat{\beta}$. The measure of the model fit is

$$RSS(\beta) = \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 = (Y - X\beta)^T(Y - X\beta) \tag{3.26}$$

where T is the notation for matrix transpose, x_i is the column vector of the predictor variables for the i^{th} observation - that is x_i is the i^{th} row of X. The estimator is

$$\widehat{\beta}_{\text{OLS}} = \arg\min_{\beta\in\mathbf{R}^p} \; RSS(\beta) = (X^TX)^{-1}X^Ty \tag{3.27}$$

When p is large, $\hat{\beta}$ has large variance, so it leads to bad results. In order to get a better extimator, we use the ridge regression method:

$$\begin{aligned}\widehat{\beta}_{\text{ridge}} \arg\min_{\beta\in\mathbf{R}^p} &\left\{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\right\} \\ &= \arg\,\min(y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta \\ &= (X^TX + \lambda I_p)^{-1}X^Ty\end{aligned} \tag{3.28}$$

where λ is a turning parameter. When $\lambda = 0$, the penalty part goes to zero, so $\hat{\beta}_{ridge} = \hat{\beta}$, i.e. the ridge regression is the same as the OLS estimator. We can find a suitable value of λ by cross-validation.

The common method to assess the goodness-of-fit of a regression model is by using the coefficient of determination $\mathbf{R}^2$, which is the ratio of the variance explained to the total variance of the dependent variable Y.

$$\begin{aligned}\mathbf{R}^2 =& 1 - \frac{RSS}{TSS} \\ &= 1 - \frac{(X\hat{\beta} - Y)^T(X\hat{\beta} - Y)}{(Y - \bar{Y})^T(Y - \bar{Y})}\end{aligned} \tag{3.29}$$

where TSS is the total sum of squares, and large values of $\mathbf{R}^2$ indicate high correlation, with 1 being the largest value and representing a perfect linear relationship without noise.

3.4.2 Prediction Results of Ridge Regression

Figure 3.2 shows the coefficient of determination in three wealth scenarios at four loadings. The first plot of figure 3.2 shows the R^2 under wealth function 1 for all

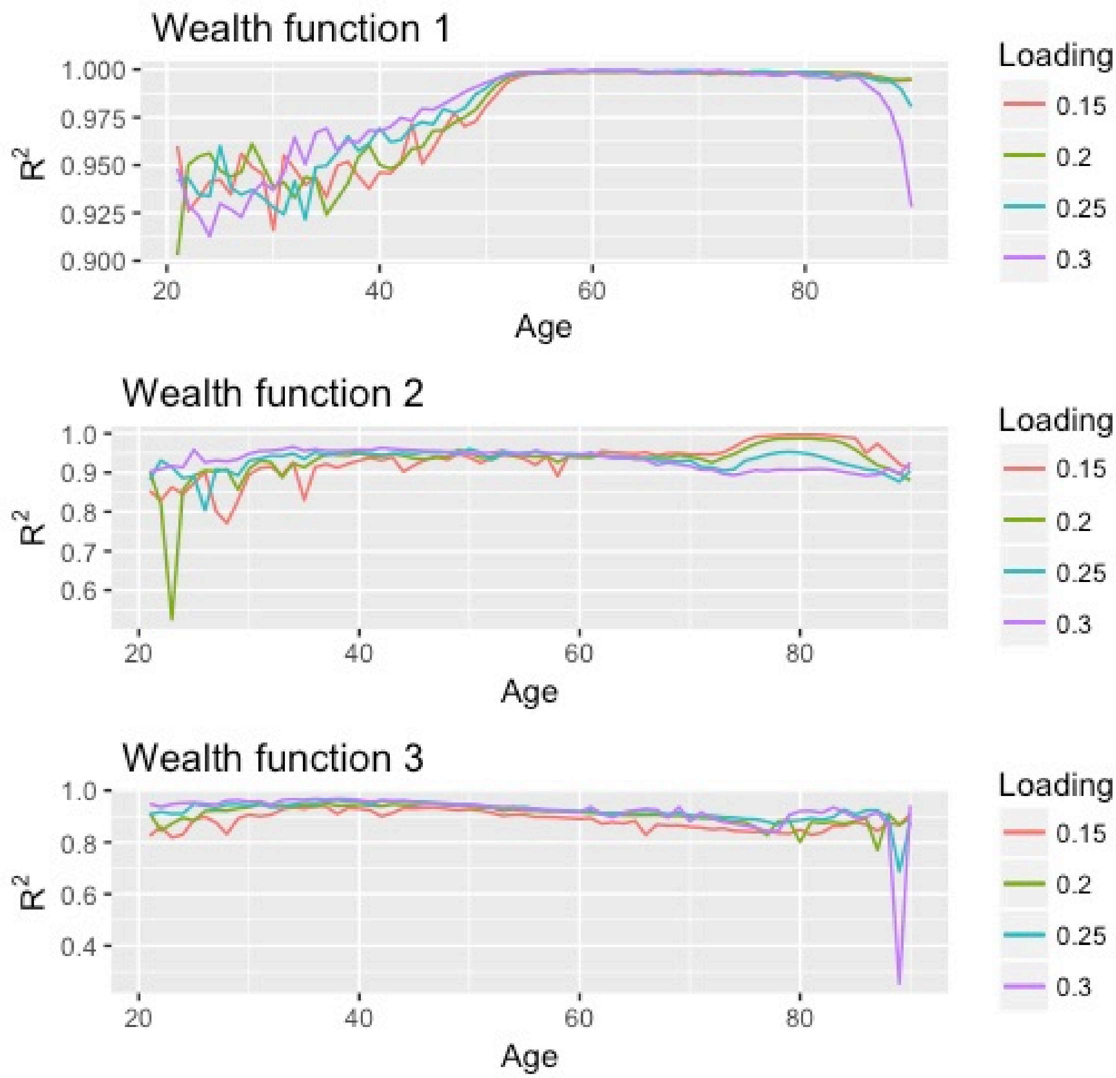

Figure 3.2: All wealth scenarios R^2 at four loadings of ridge regression.

loadings. The accuracy shows a similar pattern for all loadings. At all loadings, the model is very accurate, indicating that the chosen variables are good linear predictors of insurance choices. For wealth function 2 (second plot), the model is accurate over all ages, but slightly less accurate for loading 0.2 at age 22. For wealth function 3 (third plot), the accuracy is slightly lower at loading 0.3 at age 89,but still good.

These results give a baseline for assessing our reduced-dimension variables. If the prediction is comparable to this baseline or higher, then we have not lost much signal. If the prediction is much lower, then we have reduced the dimension too much.

3.5 Dimension Reduction

3.5.1 Singular Value Decomposition

From the ridge regression, we get a large coefficient table (699×840) containing one set of coefficients (699×70) for each scenario. We want to reduce the dimension of the predictors in order to find a more simple model of utility to predict insurance choice. Singular Value Decomposition (SVD) is a commonly used dimension reduction method that should be suitable for this problem. Given a high-dimensional (699×840) data matrix B= $(\beta_1, \beta_2, \cdots, \beta_{840})$, we applied SVD on B, where $B \approx UDV^T$. We aims to find a low-rank factorization XU where X is 1000×699, U is $699 \times r$. The idea is to make $\hat{Y} = XUDV^T$ as close to Y as possible subject to these rank restrictions. We can then replace the high-dimension X with a low-dimensional XU, meaning we need to only measure r factors of the utility function.

Our B matrix consists of coefficients fitted by ridge regression in each scenario. We use SVD to reduce the dimension of the coefficient matrix B and set the first several principal scores as our new predictors in the risk aversion dataset.

Suppose the coefficient matrix B includes 699 variables, with 840 observations for each variable, for the i^{th} variable we let $B_i = (\beta_{i1},\ \beta_{i2},\ \ldots,\ \beta_{i840})$, then the matrix is 699×840. We can write the matrix B into B= $(\beta_1, \beta_2, \cdots, \beta_{840})$, To reduce the dimension, we want to use fewer variables for prediction. For $\beta_1,\ \beta_2,\ \ldots,\ \beta_{840}$, we set those as the original variables, and we set $q_1,\ q_2,\ \ldots,\ q_r$ $(r \leq m)$ as the new variables, the principal component analysis can reduce m original variables to r variables, which can be calculated by:

$$\begin{cases} q_1 = \alpha_{11}\beta_1 + \alpha_{21}\beta_2 + \cdots + \alpha_{m1}\beta_m \\ q_2 = \alpha_{12}\beta_1 + \alpha_{22}\beta_2 + \cdots + \alpha_{m2}\beta_m \\ \vdots \\ q_r = \alpha_{1r}\beta_1 + \alpha_{2r}\beta_2 + \cdots + \alpha_{mr}\beta_m \end{cases} \tag{3.30}$$

where α_{ij} should satisfy $\alpha_{1i}^2 + \alpha_{2i}^2 + \cdots + \alpha_{mi}^2 = 1$. Equation (3.30) can be written in matrix form $q_i = B\alpha_i, (i = 1, 2, \ldots, r)$. If we select the first r principal components to restructure the matrix, we get a reduced matrix.

One approach to selecting the number of principal components is by the cumulative proportion of variance explained: the higher the proportion of explained variance, the better approximation we get.

This is shown in Figure 3.3. The first three principal components can explain almost 98% of the variance, which should be enough accuracy for our propose. Therefore, we use the first three principal score vectors as the reduced-dimension data matrix, then we fit a linear function of the predictors. Figure 3.4 shows the accuracy from using the first three principal scores of the predicted model. This accuracy is generally comparable to ridge regression, indicating that the three principal components have captured most of the signal from the original 699 variables.

3.6 Generalized Additive Model Estimation

The ridge regresison is based on the model that the amount of insurance to buy should be a linear function of the predictors. However, this might not be the case. We therefore investigate using non-linear predictiors from the principal scores. In particular, we fit a Generalized Additive Model (GAM).

A GAM has the form

$$g(E(y)) = s_1(x_1) + s_2(x_2) + \cdots + s_m(x_m) \tag{3.31}$$

where y is the dependent variable, the predictors x come from the matrix XU, $s(x)$ is a non-parametric smooth function, g is a fixed link function and $E(y)$ is the conditional expectation of y given $x_1,\ x_2,\ ...,\ x_m$.

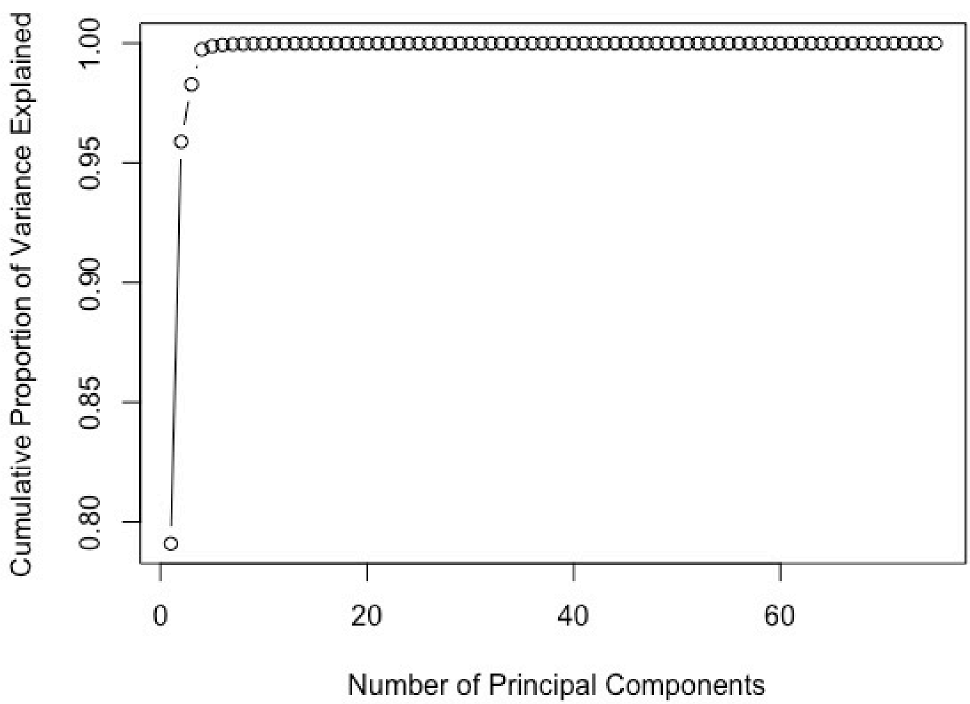

Figure 3.3: Cumulative proportion of variance explained by PCA at loading 0.15 for wealth function 1.

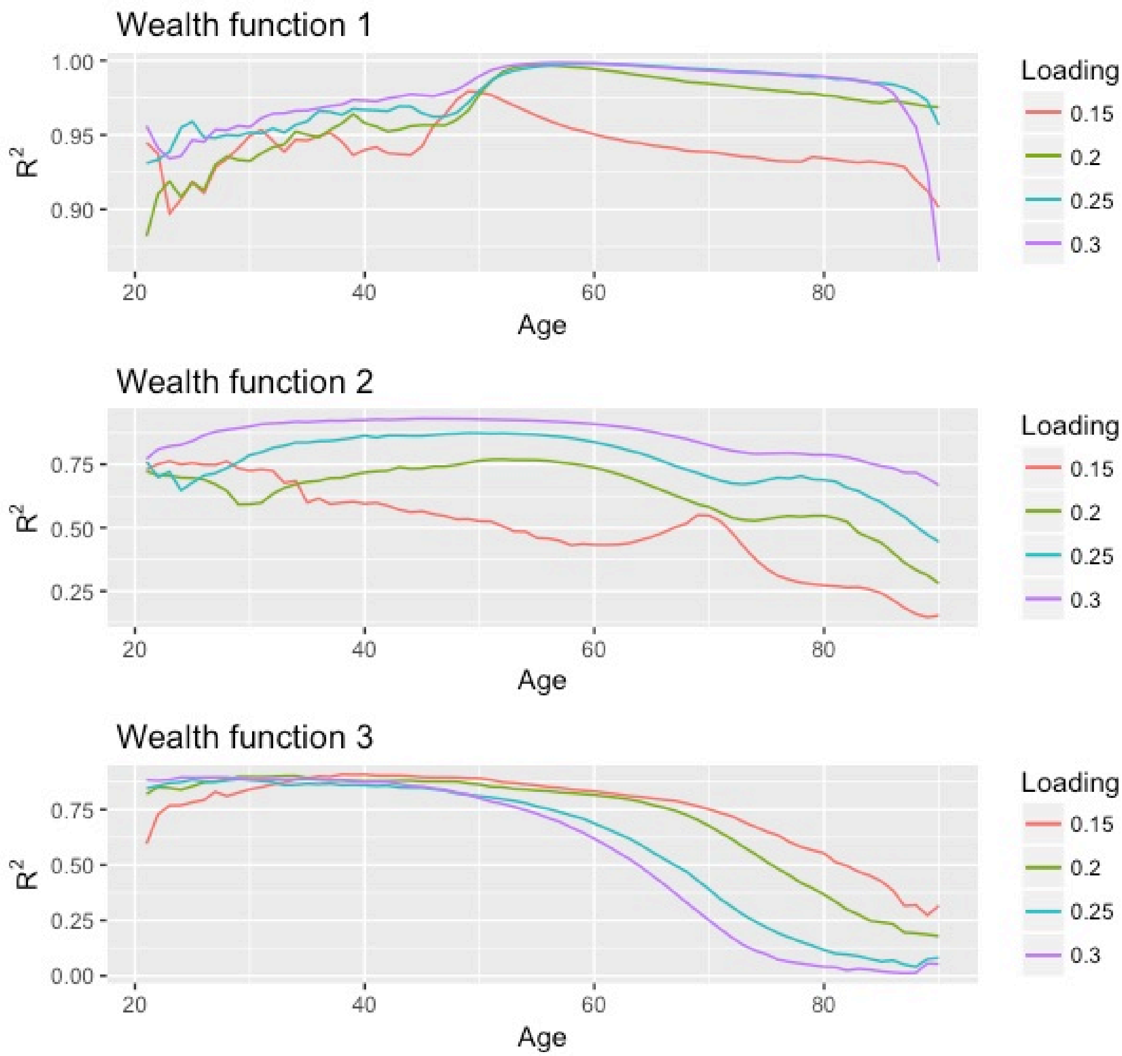

Figure 3.4: Linear model R^2 of four loadings(represented by the four colours) using PCA.

Suppose we have data matrix $B_{n\times m} = (x_1, x_2, \ldots, x_n)^T$, and we get the reduced dimension matrix from Section 3.4. In this case, we consider the first r principal scores $q_1, q_2, ..., q_r$, and use an identity link function.

Thus, the GAM on the r principal scores is:

$$y = s_1(q_1) + s_2(q_2) + \cdots + s_k(q_r) + \epsilon,\ 0 < k < m \tag{3.32}$$

We compare the R^2 values for this GAM to the R^2 values for the original ridge regression and linear model to see whether adding non-linear effects improves prediction, and by how much. The complexity of the model is controlled by the number of degrees of freedom in each s_i. We fit the GAM using the *mgcv* package which chooses the effective degrees of freedom using generalised cross-validation.

3.6.1 Results for GAM

Figure 3.5 shows the accuracy of GAM over three wealth scenarios at four loadings. Compared with Figure 3.4, we see that the accuracy using GAM is higher than using the linear model in all simulations. However, the accuracy is still low for wealth function 3.

Comparing the three wealth scenarios, we find that the prediction is best for wealth function 1 and worst for wealth function 3. There is also a large difference between prediction accuracy for different mortalities. On the other hand, the percentage loading has less effect on the accuracy of estimating the policyholder's insurance choice.

From the GAM results, the prediction of all wealth scenarios are improved. Our aim is to develop a method that works for all scenarios, hence, we want to merge all wealth scenarios. First of all, we merge the coefficients of ridge regression from each scenario, then select first k principal components. We fit a non-linear model to find the performance of the prediction. Figure 3.6 compares the accuracy of the four dimension-reduced methods at wealth function 1. (The results for wealth functions 2 and 3 are in the appendix, Figures A.1 and A.2.) From the four plots, we can see, the loadings have a small effect on the prediction. There is also a clear difference over mortalities. The generalized additive model performs better than the linear model. In many cases, the prediction is very good. However, in other cases, there is still room for improvement.

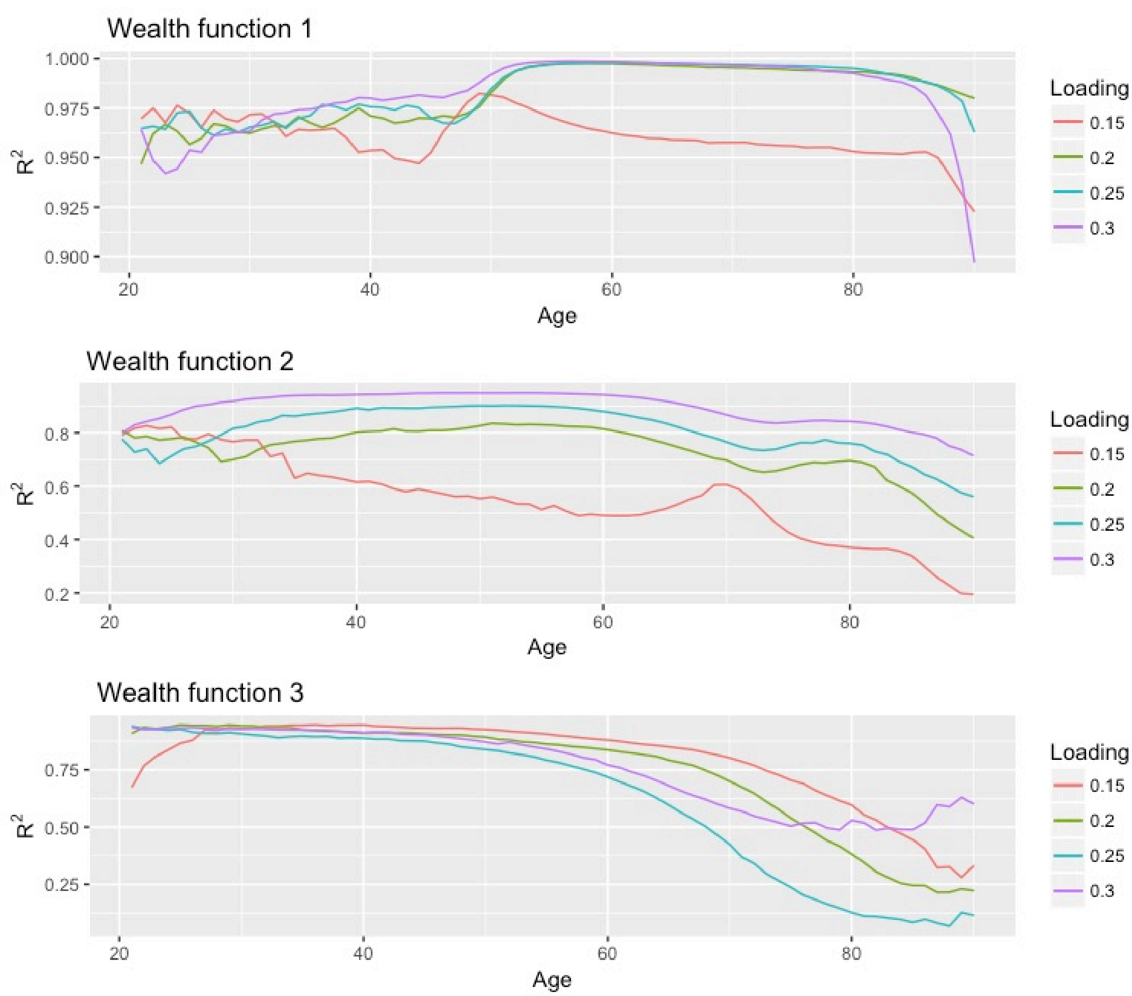

Figure 3.5: Generalized additive model R^2 for four loadings using PCA.

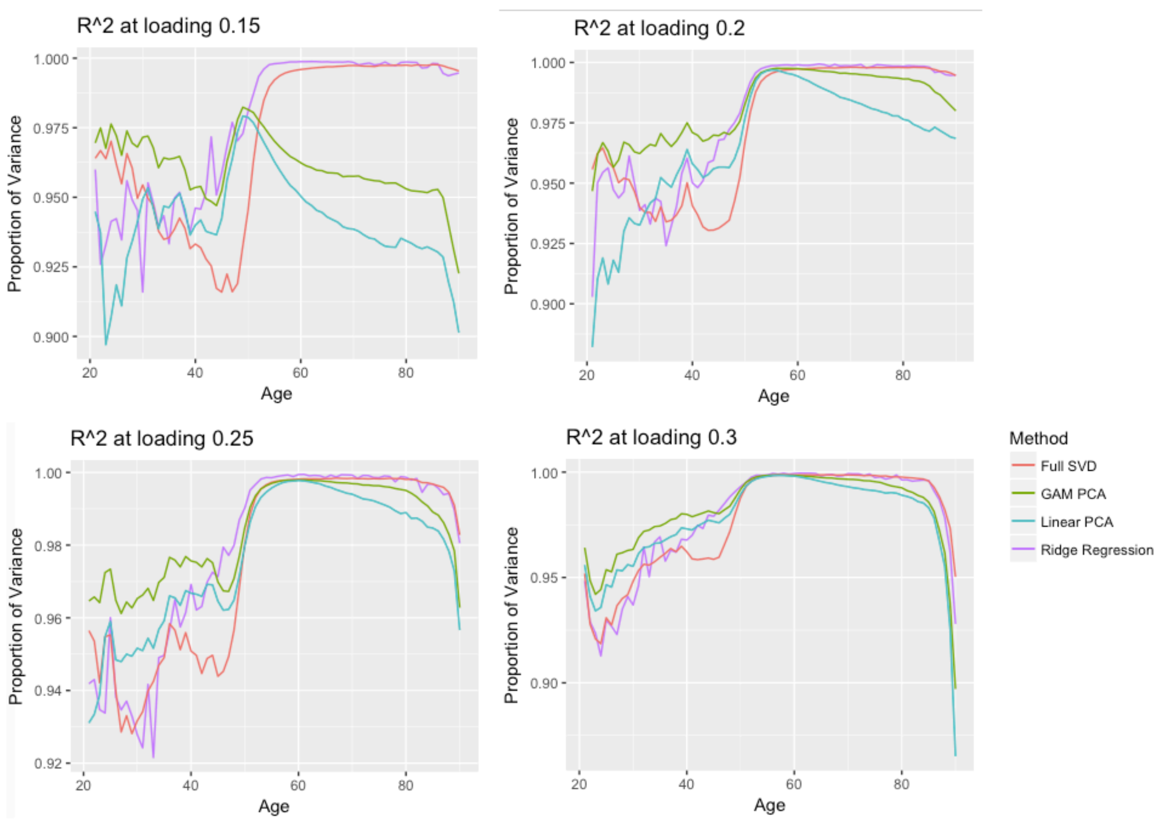

Figure 3.6: R^2 for various mortalities for All reduced-dimension methods for wealth function 1.

Chapter 4

Describing the Principal Components.

We aim to estimate a utility function model which can give suggestions to the potential policyholder about the insurance choices. However, it is difficult to estimate the utility function. From Chapter 3, we found that estimating a small number of features of the utility function can give a good estimate of the individual's insurance preferences. In this chapter, we seek to describe the features identified.

4.1 Polynomial Features

In the Chapter 3, we computed a 699-dimensional description of the utility functions by evaluating ARA at 699 values. Based on the results of the final model, we found that 3-7 linear combinations of these variables can get a reasonable model to predict the insurance choice. In the full predicted model, when we use the first three principal scores, the predicted model accuracy is acceptable. However, we would like to develop a better description of these three principal components, so that future work on measuring these features of the utility function is feasible.

Figure 4.1 shows the first three principal components. We aim to get a good fit of wealth range between 0 and 2.17. From Figure 4.2, visually, it looks as if these functions can be roughly approximated by polynomial functions. We attempt to fit polynomials to these functions and use the polynomial approximations to construct features. We then fit a GAM on these polynomial features to predict insurance choices.

Figure 4.3 shows the accuracy when we replace the principal component vectors by polynomial approximations. We see that for wealth function 1 (The results for wealth functions 2 and 3 are in the appendix, Figures A.3 and A.4.), the polynomial approximation retains most of the prediction accuracy of the original principal components, even though the approximation is less accurate for lower ages, the accuracy is acceptable, indicating that estimating a low-order polynomial-weighted average of

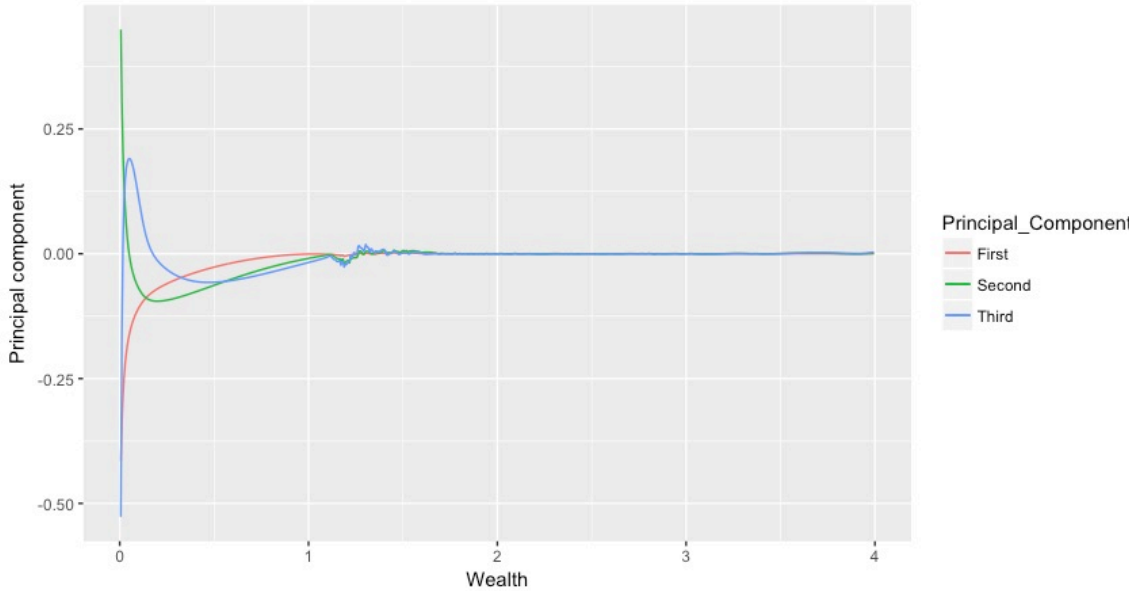

Figure 4.1: First three principal components of different wealth values

risk-aversion may provide a good estimator for the individual's insurance choice. The approximated principal components give higher accuracy at older ages and lower accuracy at younger ages. We also attempt to use more principal components to catch more signal. We compare the performance at ages 25,45, 65 and 85, to indicate the model behaviors. In practice, the loading has less effect on the accuracy, we therefore fix the loading at 0.15. Figure 4.4 shows the R^2 at those ages. Based on this, we decide to use the first six principal components.

Table 4.1 and Table 4.2 measure the R^2 based on the polynomial approximation for wealth function 3. Comparing two tables, as the number of principal components increases, the accuracy improves. When we use six principal components, the accuracy is fairly acceptable. Figure 4.5 shows the comparison between all reduced-dimension methods and polynomial approximation by using the first three and six principal components. For wealth function 3, we do better by increasing the number of principal components. Comparing the figures and tables, these indicate that using polynomials works fairly well for all three wealth functions.

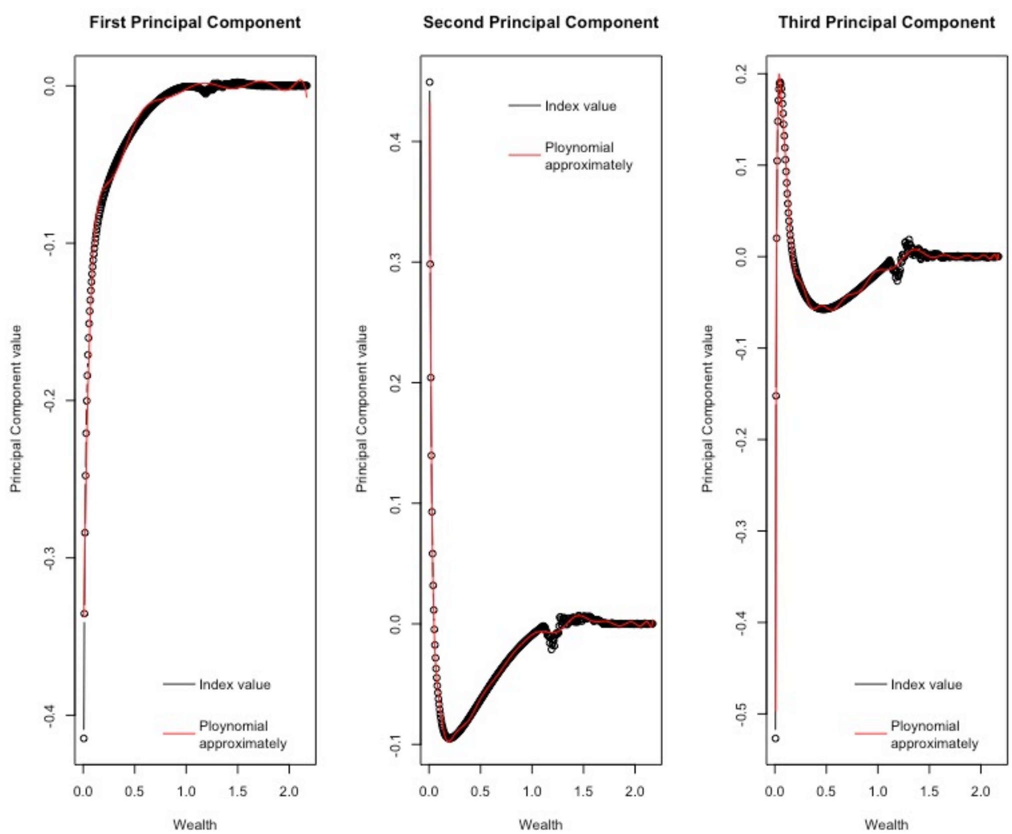

Figure 4.2: Polynomial approximation of first three principal components.

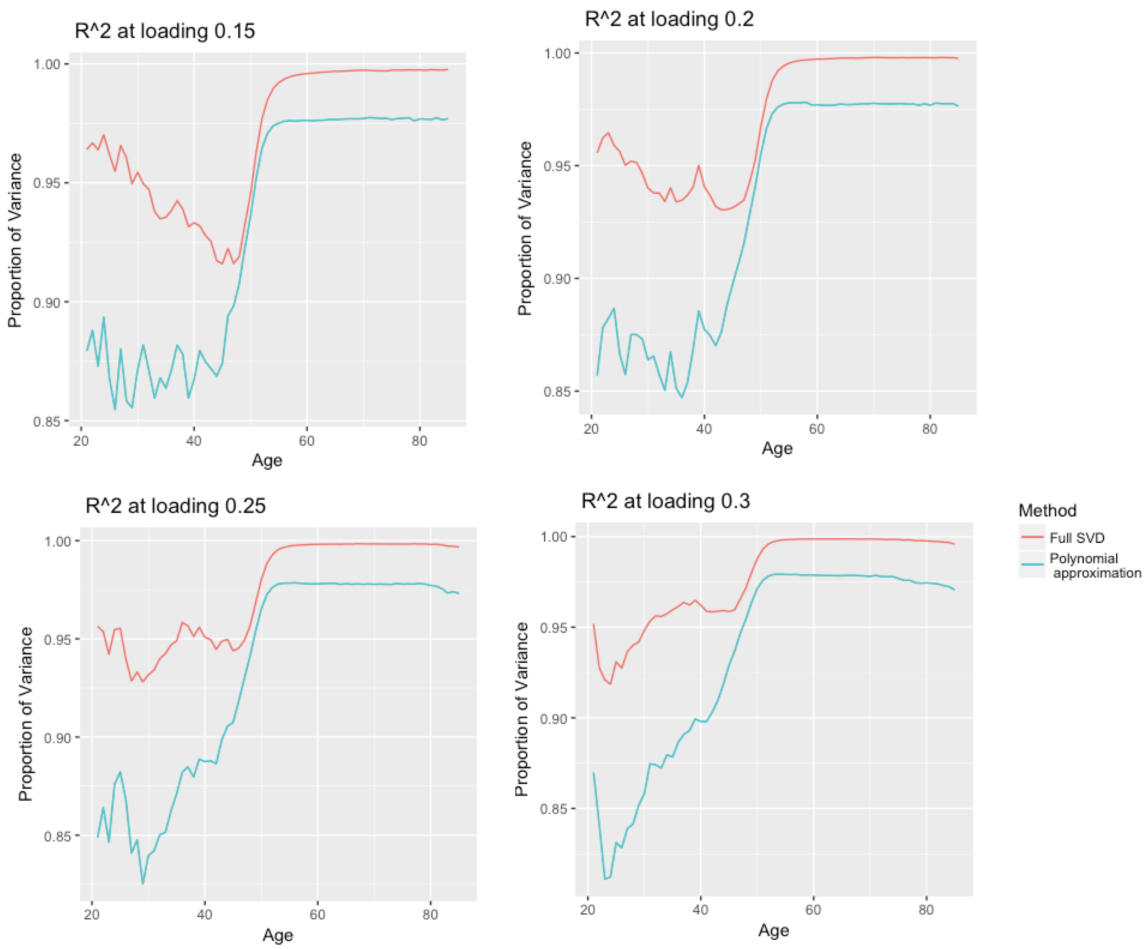

Figure 4.3: R^2 for various mortalities for Polynomial approach for wealth function 1.

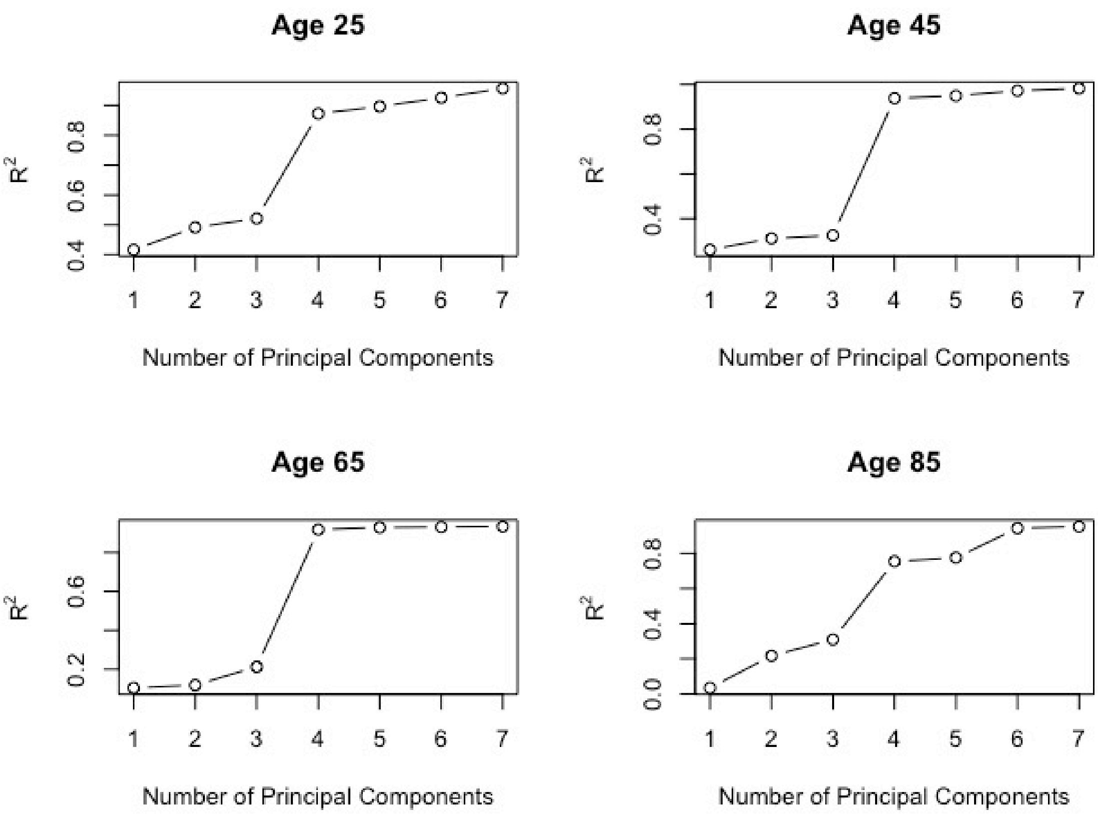

Figure 4.4: R^2 of models using different Numbers of principal components with polynomial approximation approach at four specific ages of loading 0.15 under wealth function 3.

Table 4.1: Polynomial Approach on first three principal components R^2 for GAM prediction for wealth function 3.

Loading	0.15	0.2	0.25	0.3
Age 5	0.5213537	0.4389122	0.4394909	0.4389122
Age 25	0.3254057	0.3010330	0.2610119	0.3010330
Age 45	0.2123227	0.1955919	0.1618593	0.1955919
Age 65	0.3097899	0.3232740	0.3444185	0.3232740

Table 4.2: Polynomial Approach on first six principal components R^2 for GAM prediction for wealth function 3.

Loading	0.15	0.2	0.25	0.3
Age 25	0.9254903	0.9487639	0.9662105	0.9710329
Age 45	0.9720814	0.9734580	0.9760268	0.9756148
Age 65	0.9315392	0.9372890	0.9418986	0.9480444
Age 85	0.9434526	0.9490275	0.9675529	0.9700478

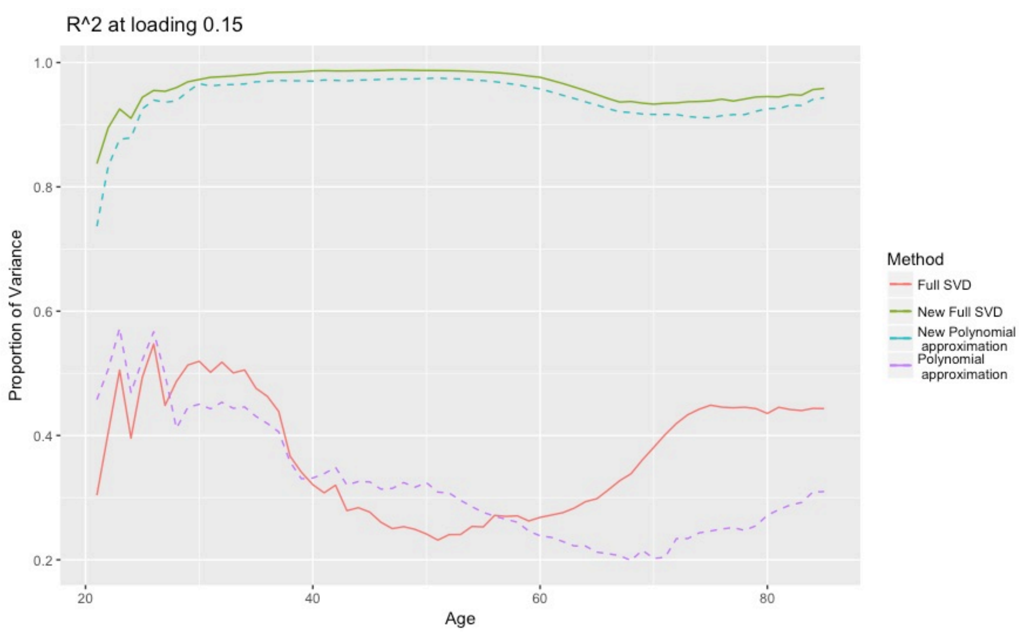

Figure 4.5: R^2 for various mortalities for Polynomial approach at loading 0.15 for wealth function 3. (The red solid line represents the non-linear model performance by using first three principal components of the merged matrix. The green solid line represents the non-linear model performance by using first six principal components of the merged matrix. The blue dashed lines represents the polynomial approximation by using first six principal components. The purple dashed lines represents the polynomial approximation by using first three principal components.)

Chapter 5

Conclusion and Further work

5.1 Conclusion

In this thesis, we have investigated the effect of utility function on insurance choices. We found that for common parametric utility functions, the choice of funcrion has a very large effect on insurance choice.

We then performed a large-scale simulation with random utility functions across a range of insurance choices to identify key features of the utility function for predicting an individual's insurance choices. We found that in many cases, an individual's insurance choices could be well predicted basd on just three features of the utility function. For a few cases, we need to increase this to six features. Furthermore, we can approximately measure these features as polynomial-weighted averages of the individual's risk aversion. While the results are encouraging, there were some cases where the prediction accuracy was limited, such as for exponentially increasing wealth. There are a number of possible reasons why this might be the case, for example: in some cases, the range of insurance choices is too narrow, so when we do the prediction, the noise is relatively large.

5.2 Further work

Further work is needed to ensure the accuracy can be improved, and to test that the dimension-reduction approach works over a wider range of scenarios, like the wealth following an exponential pattern. Once the important features of the utility functions have been identified, more work is needed on estimating these features. This could be done by using a parametric approach like the common choices of parameterized forms as we used but with additional parameters. Alternatively, non-parametric estimation could be used. We could approximate features in different ways, such as using rational functions instead of polynomials.

Appendix A

Figures of other wealth scenarios

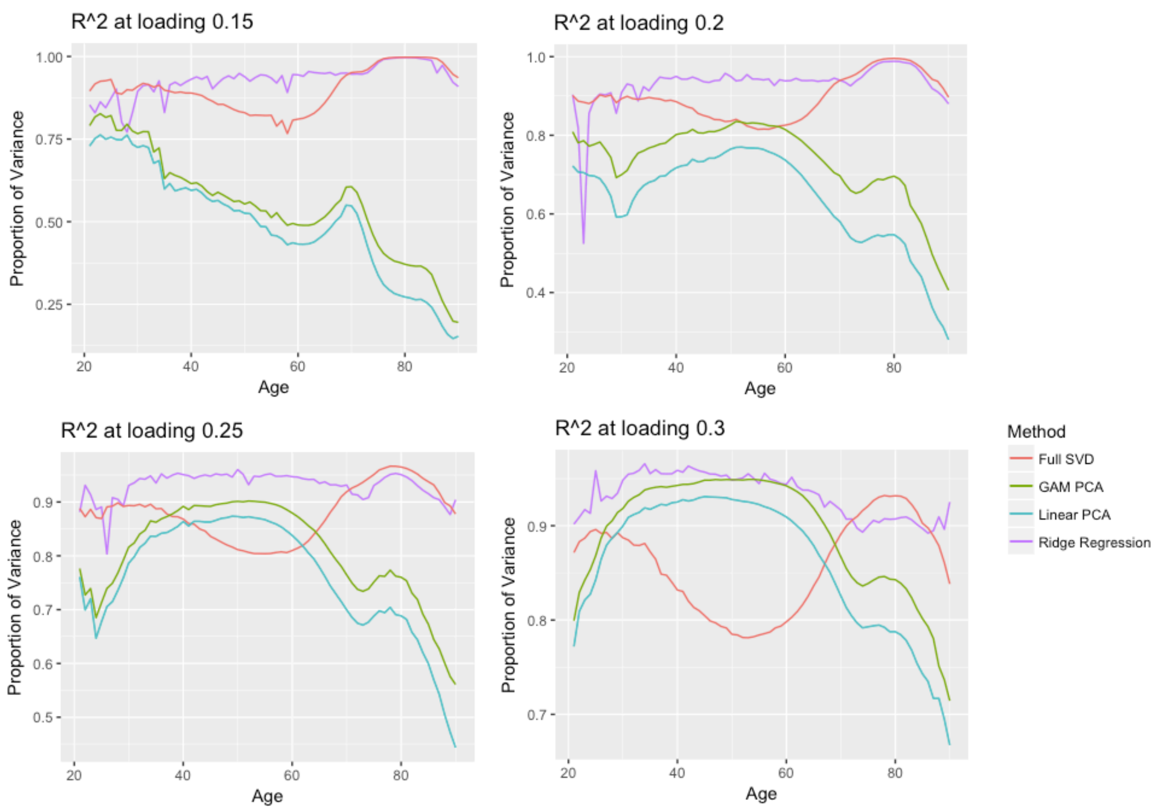

Figure A.1: R^2 for various mortalities for All reduced-dimension method for wealth function 2.(The red line represents the non-linear model performance of the merged matrix.)

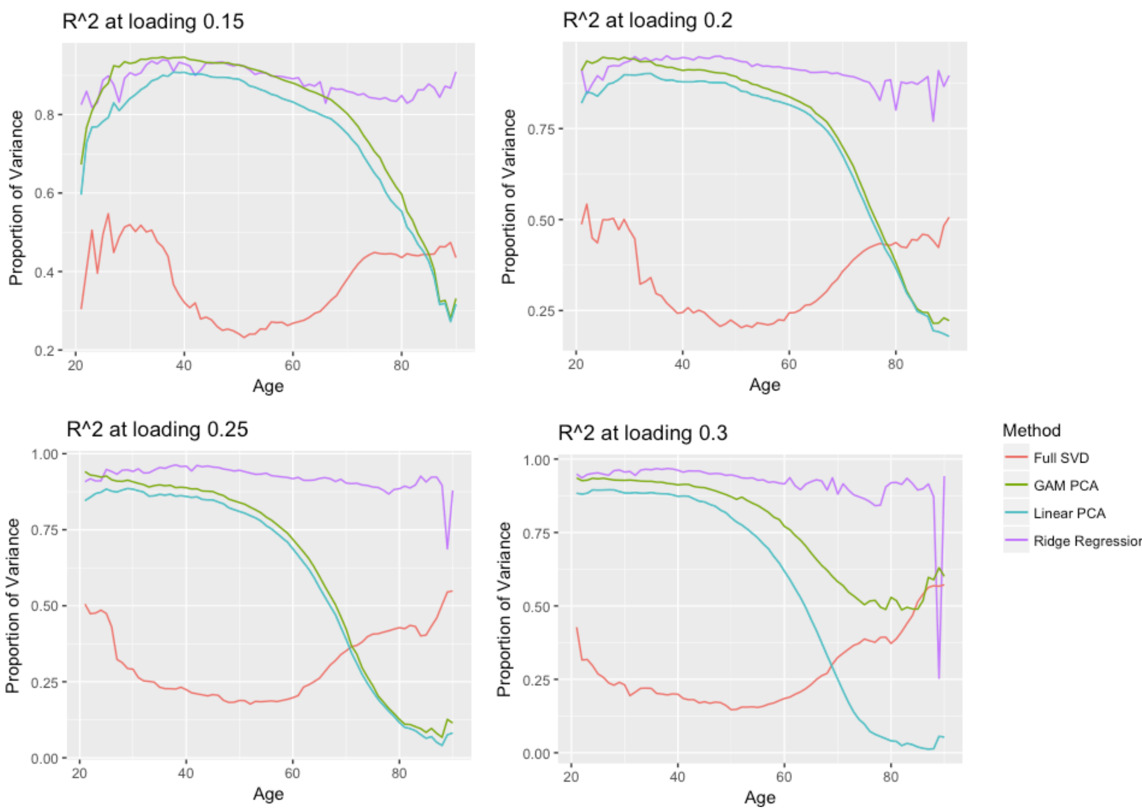

Figure A.2: R^2 for various mortalities for All reduced-dimension method for wealth function 3.(The red line represents the non-linear model performance of the merged matrix.)

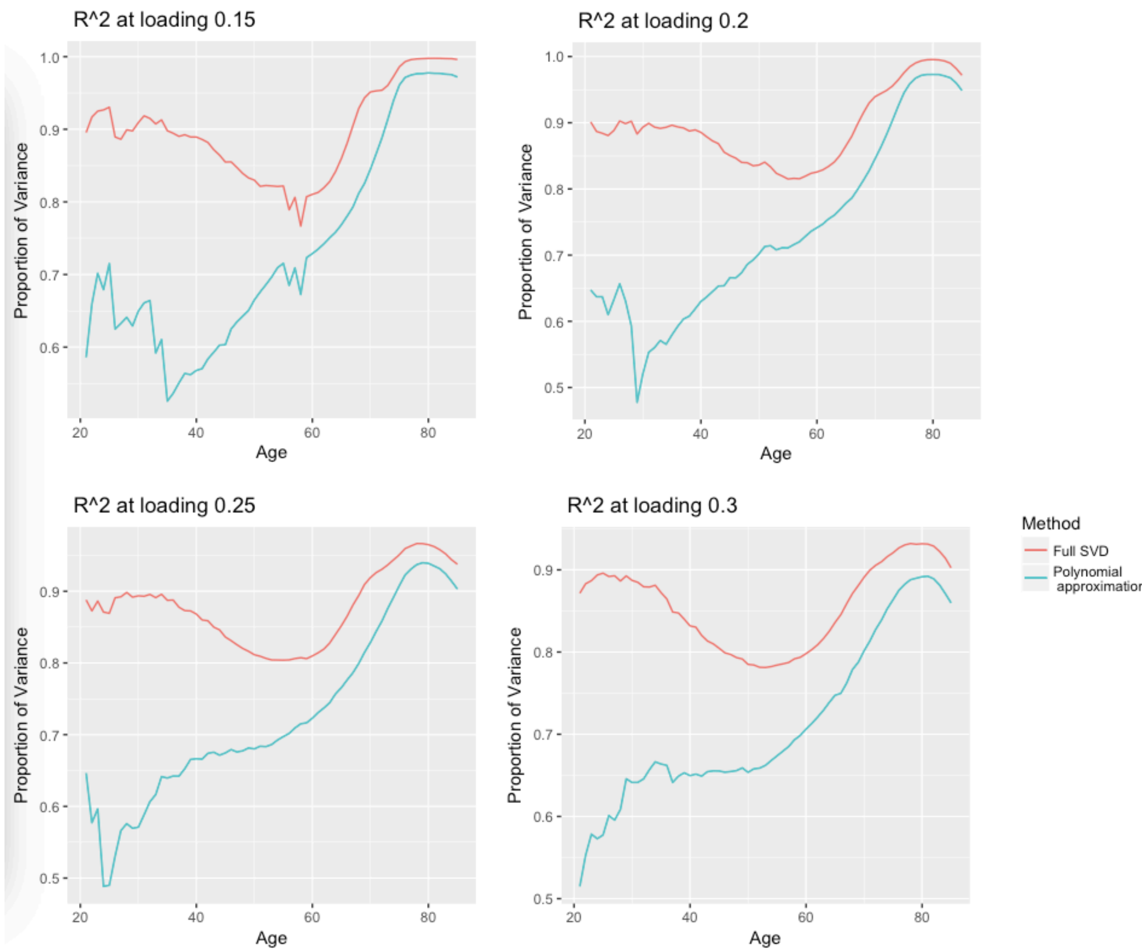

Figure A.3: R^2 for various mortalities for Polynomial approach for wealth function 2.

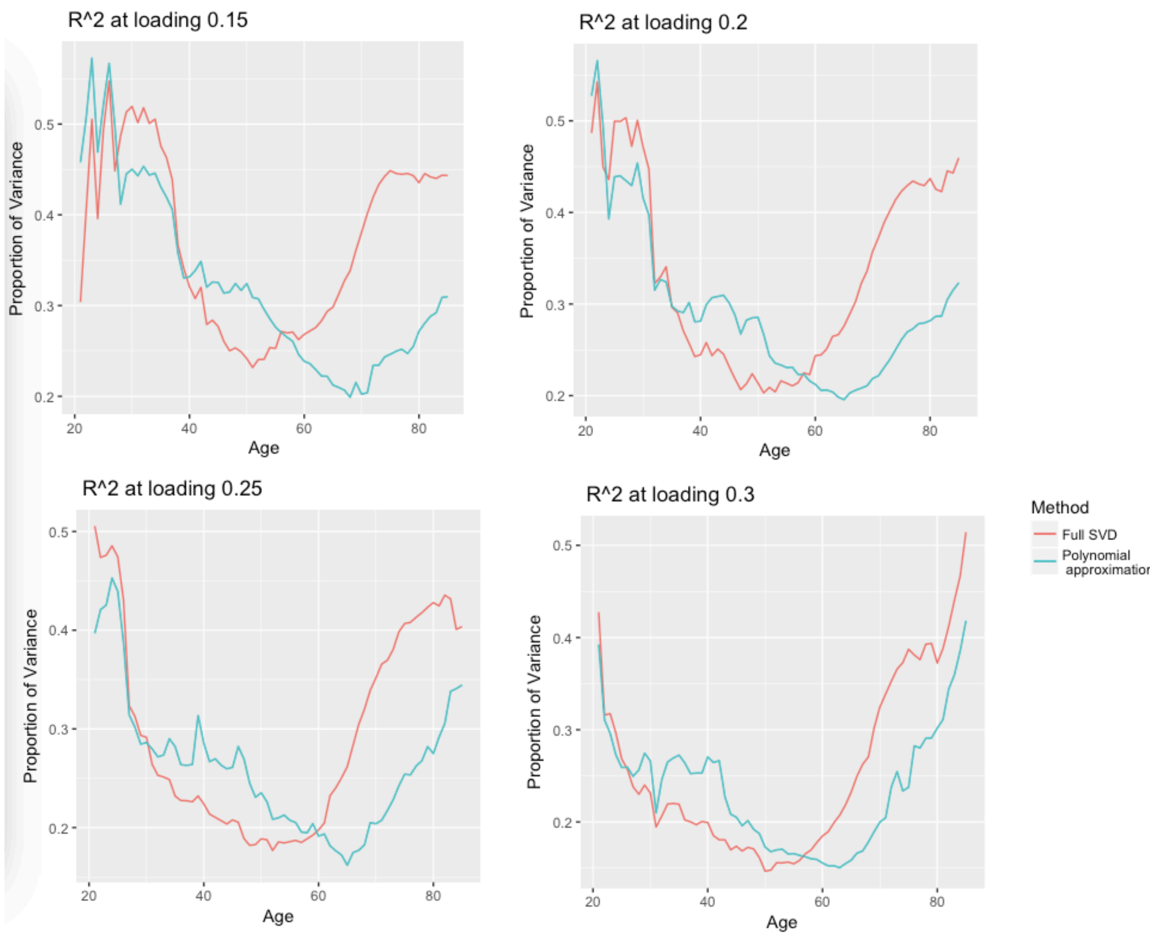

Figure A.4: R^2 for various mortalities for Polynomial approach for wealth function 3.

Bibliography

[1] David Blake. Efficiency, risk aversion and portfolio insurance: an analysis of financial asset portfolios held by investors in the united kingdom. *The Economic Journal*, pages 1175–1192, 1996.

[2] Patrick L Brockett and Linda L Golden. A class of utility functions containing all the common utility functions. *Management Science*, 33(8):955–964, 1987.

[3] Gianni Cicia, Teresa Del Giudice, and Riccardo Scarpa. Consumers perception of quality in organic food: a random utility model under preference heterogeneity and choice correlation from rank-orderings. *British Food Journal*, 104(3/4/5):200–213, 2002.

[4] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[5] Barry K Goodwin and Vincent H Smith. *The economics of crop insurance and disaster aid*. American Enterprise Institute, 1995.

[6] William M Gorman. The structure of utility functions. *The Review of Economic Studies*, 35(4):367–390, 1968.

[7] Martin F Grace and Stephen G Timme. An examination of cost economies in the united states life insurance industry. *Journal of Risk and Insurance*, pages 72–103, 1992.

[8] Ludger Hinners-Tobrägel. The st. petersburg paradox. In *EFIIA 2003 Conference Procedings*, pages 867–875, 2003.

[9] Arthur E Hoerl, Robert W Kannard, and Kent F Baldwin. Ridge regression: some simulations. *Communications in Statistics-Theory and Methods*, 4(2):105–123, 1975.

[10] TC Johnson. Utility functions. *C2922 economics, Heriot Watt University, Edinburgh Google Scholar*, 2007.

[11] James M Joyce. St. petersburg paradox. In *International Encyclopedia of Statistical Science*, pages 1377–1378. Springer, 2011.

[12] Marek Kuczma. *An introduction to the theory of functional equations and inequalities: Cauchy's equation and Jensen's inequality.* Springer Science & Business Media, 2009.

[13] Donald W Marquaridt. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612, 1970.

[14] Policonomics. Risk and uncertainty i: Risk aversion. `http://policonomics.com/lp-risk-and-uncertainty1-risk-aversion/`. Accessed February, 2016.

[15] Murray Newton Rothbard. Toward a reconstruction of utility and welfare economics, 1956.

[16] Lloyd S Shapley. The st. petersburg paradox: A con games? *Journal of Economic Theory*, 14(2):439–442, 1977.

[17] Giorgio Alfredo Spedicato et al. The lifecontingencies package: Performing financial and actuarial mathematics calculations in r. *Journal of Statistical Software*, 55(10):1–36, 2013.

[18] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[19] Walter B Weare. *Black Business in the New South: A Social History of the NC Mutual Life Insurance Company*. Duke University Press, 1993.