OPTIMAL HEALTH INSURANCE WITH CONSTRAINTS UNDER
UTILITY OF HEALTH, WEALTH AND INCOME

YAN ZHANG
School of Finance, Guangdong University of Foreign Studies
Guangzhou 510006, China

YONGHONG WU
Department of Mathematics and Statistics, Curtin University
Western Australia 6845, Australia

HAIXIANG YAO*
School of Finance, Guangdong University of Foreign Studies
Guangzhou 510006, China

(Communicated by Ka Chun Cheung)

Abstract. We consider an optimal health insurance design problem with constraints under utility of health, wealth and income. The preference framework we establish herein describes the trade-off among health, wealth and income explicitly, which is beneficial to distinguish health insurance design from other nonlife insurance designs. Moreover, the work takes into account the case that if the insured is severely or critically ill, the insured may not fully recover even after necessary medical treatment. By taking these special features into account, the health insurance design problem is formulated as a constrained optimization problem, and the optimal solutions are derived by using the Lagrange multiplier method and optimal control technique. Finally, two numerical examples are given to illustrate our results. Our research work gives new insights into health insurance design.

1. Introduction. In order to promote individual and population health and improve health care system, the National People’s Congress of China set out the “Healthy China 2030” plan in 2016, as part of the comprehensive strategy for building a more healthy China (Wu et al. [33]). It is an innovative development idea of health priority, and a flag combining the common ideal government, society and all people (Li and Wang, [21]). In this paper, we formulate the health insurance design as an optimization problem and derive the corresponding solution under a more realistic utility function.

The optimal insurance design problem is one of the most important problems in actuarial science. Pioneered by Arrow [1, 2, 3], many articles on optimal insurance design began to proliferate. For example, Arrow [1] developed the framework of expected utility maximization, which has been widely adopted in the field of economics and finance.

2020 Mathematics Subject Classification. Primary: 91B30, 97M30; Secondary: 46N10, 47N10.

Key words and phrases. Optimal health insurance, expected utility, shocks to health status, wealth, income, indemnity, deductible.

This research work is supported by the National Natural Science Foundation of China (Nos. 12001119, 71871071, 72071051), the Natural Science Foundation of Guangdong Province of China (Nos. 2017A030313399, 2018B030311004), and the Featured Innovation Project (Youth Foundation) of Guangdong University of Foreign Studies (19QN21).

* Corresponding author: Haixiang Yao.
problem under various constraints, with consideration of application in personal lines or commercial lines, have been contributed to the literature in this field. Typical papers include but are not limited to Smith [31], Mossin [25], Spence and Zeckhauser [32], Raviv [30], Doherty and Schlesinger [8], Gollier and Schlesinger [16], Young [37], Moore and Young [24], Lee [19], Zhou and Wu [42], Zhou et al. [43], Bernard et al. [4], Promislow and Young [28], Lu et al. [22], Xu et al. [35], Ho et al. [18], Fan [11] and Zhang and Wu [40]. Most of the aforementioned papers assume that the individual would suffer a random loss and pay premium to the insurer for indemnity in case that the loss occurs. Under Arrow’s framework, the optimal insurance is deductible insurance under both static setting and dynamic setting. Some general qualitative properties of the optimal deductible have been obtained under different utility of wealth, for example, exponential utility, power utility, logarithmic utility, IARA utility, and DARA utility.

The conclusion and research results in most existing literatures are derived based on utility of wealth only, which appears unrealistic in the sense that insurance premium is in practice paid out of one’s income, rather than from wealth or property (Lee, [19]). Especially, premium is usually paid monthly or quarterly for short-term health insurance. Therefore, it is necessary to distinguish wealth from income in utility function. Moreover, as a special type of nonlife insurance, the study on optimal health insurance contract design should take the insured’s health status into account. Hence, the first important problem in our study is to design a realistic utility function to describe the interactions among the three key factors, including health $h$, wealth $w$, and income $y$. In the past decades, numerous researchers have investigated utility of health (Fuchs and Zeckhauser, [14]; Pliskin et al. [27]), utility of wealth (Markowitz, [23]), trade-off between health and wealth (Levy and Nir, [20]; Finkelstein and Luttmer, [12]), health state-dependent utility functions that describe job injury and impact on income (Viscusi and Evans, [34]). Inspired by previous research works, we apply the Time Trade-off (TTO) approach to derive a utility that is linear in health. We also utilize the idea from Viscusi and Evans [34], Lee [19], Levy and Nir [20] to propose an explicit utility function taking into account not only health and wealth but also income.

With the utility function established, we then formulate the health insurance problem into a mathematical model. By the Lagrange multiplier method and optimal control technique, we obtain the optimal indemnity. Compared with other research articles on health insurance, for example, Zeckhauser [38], Phelps [26], Besley [5], Blomqvist [7], Hall and Jones [17], Ellis et al. [10], Bhargava et al. [6], Yan [36], Finkelstein et al. [13], Gerfin [15], Zheng et al. [41], and Doiron and Kettlewell [9], our research focuses on optimal indemnity design, instead of discussion about taxation effect, demand for health insurance or health care, health spending, adverse selection, or moral hazard.

The contribution of our paper is threefold. First, we formulate the utility and optimization problem mathematically by considering the interactions among health, wealth, and income. Second, we derive the optimal solutions by applying the Lagrange multiplier method and optimal control technique. Third, we illustrate the optimal indemnity contract by comparison of results under different loss distributions. We gain new insights into health insurance design. To summarize, the main results of our work are as follows.

1. For illustration purpose, the optimal indemnity is obtained under Exponential and Pareto loss distributions, respectively.
2. There is a critical value of income $y$ and upper limit on coverage $K$. It means that the optimal insurance turns from partial coverage insurance to full coverage insurance when $y$ or $K$ varies over the interval including critical value.

3. There is no sudden switch from full to partial coverage insurance due to the slight increase of safety loading factor from zero.

4. The current health status $h$, current wealth $w$, and degree of health degeneration $\theta$ have relatively smaller impact on the optimal indemnity.

The rest of the paper is organized as follows. In Section 2, the health insurance design problem is formulated as an constrained optimization problem with a new utility function taking into account health and wealth as well as income. In section 3, the solution of the optimization problem is obtained by using the Lagrange multiplier method and optimal control technique. Section 4 is devoted to illustration of the theoretical results by two numerical examples. Section 5 concludes the paper.

2. The Model for Health Insurance Design. In this section, we firstly establish a new utility function for health insurance taking into account insured’s health, wealth and income. Then the health insurance design problem is formulated as a constrained optimization problem for the insured to maximize the expected utility under the expected value actuarial pricing principle.

2.1. Utility of Health, Wealth and Income. In the field of health economics, the utility function measures the overall well-being of an insured as a function of various variables such as health status and wealth. It gives a criterion or objective function for the decision making in health insurance design. More specifically, with the utility function, the optimal insurance design is to be achieved in the sense of maximizing the utility function.

The trade-off between health and wealth is an important central issue for social policy and household risk management. Over the last couple of decades, a great deal of research has been conducted to study this problem, and various forms of utility functions have been proposed. For more details, the reader is referred to Viscusi and Evans [34], Hall and Jones [17]; Levy and Nir [20]; Pliskin et al. [27].

Levy and Nir [20] compared various forms of utility functions, including

- Logarithmic utility: $U(h, w) = h \cdot \log(aw)$;
- Power utility: $U(h, w) = h \cdot \frac{(w+A)^{1-a}}{1-a}$;
- Negative exponential utility: $U(h, w) = h \cdot (B - e^{-bw})$;

where $h$ represents health status, $w$ denotes wealth, $a$, $b$ and $B$ are scalar constants, $-A$ is the minimal consumption level required for existence. Analysis of the health-wealth tradeoff choices reveals that the logarithmic utility function gives the best description of choices not only at the aggregate level, but also at the individual level.

The shortcoming of the above-mentioned utility functions is that they do not distinguish between wealth and income. As the insured’s general well-being depends on health, wealth and income, it is natural to argue that income should be taken into account in the utility function. To formulate the utility of income, we refer to Viscusi and Evans [20] and Lee [19], in which wage-risk tradeoffs are considered. Viscusi and Evans [34] proposed a logarithmic utility function of income and developed a health state-dependent utility framework for the health economics and social insurance. Lee [19] dealt with the optimal insurance problem and argued that
insurance premium is in practice paid out of the insured’s income, rather than from wealth or property. To reflect this fact, the utility function should depend on both income and wealth, instead of on wealth only.

Hence in this paper, we extend the utility of health and wealth \( U(h, w) \) to the utility of health, wealth and income \( U(h, w, y) \), where \( y \) denotes income. The specific form of the utility function can be deduced through the following steps.

**Step 1:** As both the Standard Gamble (SG) approach and the Time Trade-Off (TTO) approach imply that \( U(h, w, y) \) must be linear in \( h \), we can assume that the insured’s preferences are given by \( U(h, w, y) = h \cdot U(w, y) \), where \( U(w, y) \) is increasing, strictly concave, twice differentiable with respect to \( w \) and \( y \), and \( \frac{\partial^2 u}{\partial w \partial y} = \frac{\partial^2 u}{\partial y \partial w} = 0 \). Since the derivation process is very similar to that in Levy and Nir [20], we omit the details here.

**Step 2:** Based on the research results of Levy and Nir [20] and Viscusi and Evans [34], we choose the logarithmic utility function for wealth \( \log(w) \) and logarithmic utility function for income \( \log(y) \).

**Step 3:** To reflect the time value of money, we follow the idea in Lee [19]. As the insured’s general well-being depends on health, wealth and also income, we assume \( U(w, y) = U(y) + \delta V(w) \), where \( \delta \) is a discount factor or weighting factor. Hence together with the analysis in steps 1 and 2, we get

\[
U(h, w, y) = h \cdot U(w, y) = h \cdot (U(y) + \delta V(w)) = h \cdot (\log(y) + \delta \log(w)),
\]

where \( 0 \leq h \leq 1 \), with 0 representing death and 1 representing perfect health; \( w > 0 \), \( y > 0 \). In the current insurance coverage term, the insured pays premium. In the next period, the insured’s overall utility depends on whether the wealth is reduced by illness in the previous period.

### 2.2. The Health Insurance Optimization Problem

Suppose that an insured or policyholder, with health status \( h \), wealth \( w \) and income \( y \), faces potential shocks to health status which may lead to a nonnegative continuous random financial loss \( X \) defined in the probability space \((\Omega, \mathcal{F}, \mathbb{P})\). The individual then purchases health insurance against the potential loss by paying a premium \( \pi \) to an insurer for a so-called indemnity \( I(X) \), where \( I(X) \) is assumed to be a non-decreasing function of \( x \) with \( I(0) = 0, 0 \leq I(x) \leq x, \forall x \geq 0 \). We also assume that each individual is covered by a single plan and receive health service and medical treatment if the individual is sick or injured. The insurance premium \( \pi \) is assumed to be a function of the expected benefit received by the insured and satisfies that \( (1 + \rho)E[I(X)] \leq \pi \), with \( \rho > 0 \) denoting the loading factor. To make sure that the insured has income \( y \) enough to purchase health insurance, we assume \( 0 < \pi < y \).

As Blomqvist [7], with health insurance in force, if shocks to health status occur, the insured will take medical treatment in hospital and get reimbursement for the cost and related loss from the insurer. Hence, during the term of health insurance, there are three possible cases: (i) the insured does not get any disease, and thus no claim would be made accordingly; (ii) the insured gets mild diseases and fully recovers by brief therapy, and thus the corresponding medical spending would be claimed; (iii) the insured gets severe or critical illness and the long-term treatment does not cure the insured completely (in other words, it is very common that the
insured still suffers health degeneration after treatment), and thus both the medical expenditure and related cost would be reimbursed by the insurer. Under case (iii), we assume that the insured’s income would decrease to $\alpha y$, $0 < \alpha < 1$ due to the insured’s absence from work.

Let $\Theta$ be a discrete random variable measuring the difference between the original health status before shocks occur and the health status after medical treatment ends. Suppose that $\Theta$ takes value in $[\theta_l, \theta_u]$, $0 < \theta_l < \theta_u \leq h$. For convenience, given that shocks to the insured’s health status occur, we define the related probabilities as follows:

- $p_1 = \text{Prob}[\Theta = 0]$, i.e. the probability that case (ii) occurs;
- $p_{\theta_i} = \text{Prob}[\Theta = \theta_i]$, and $p_2 = \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i}$, i.e. the probability that case (iii) occurs.

Then, the insured’s expected utility can be calculated as follows:

$$g(I(X), \pi) = E[U(X, \Theta, I(X), \pi)] = (1 - p_1 - p_2)U(h, w, y - \pi)$$
$$+ p_1 \int_{0}^{\infty} U(h, w - x + I(x), y - \pi) dF(x|\Theta = 0)$$
$$+ \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} \int_{0}^{\infty} U(h - \theta_i, w - x + I(x), \alpha y - \pi) dF(x|\Theta = \theta_i),$$

where $F(x|\Theta = \theta_i)$ is the conditional distribution of $X$ given $(\Theta = \theta_i)$, provided that shocks to the insured’s health status occur.

Under the constraint that the insurer places an upper limit $K$ on coverage, the risk-averse insured aims to maximize the expected utility against unpredictable shocks to health status. Let $I$ denote the set of indemnity designs that satisfy the conditions $I(X) \leq K$ and $0 \leq I(X) \leq X$. Let $\Pi$ be the collection of $\pi$ that satisfies the condition $\pi \geq (1 + \rho)E[I(X)]$. It means that the minimum price that could be accepted by the risk-neutral insurer is $(1 + \rho)E[I(X)]$. Then, with $I(X)$ and $\pi$ being decision variables, the optimal health insurance model can be stated as follows:

**Problem P_1.**

$$\max_{I, \pi} g(I(X), \pi)$$
$$\text{s.t.} \quad \pi \geq (1 + \rho)E[I(X)],$$
$$I(X) \leq K,$$
$$0 \leq I(X) \leq X.$$

3. **Solution Scheme.** In this section, we apply the Lagrange multiplier method together with the optimal control technique, to solve the optimization problem in two steps. In the first subsection, we keep the insurance premium $\pi$ fixed, and solve the optimal health insurance problem $I^*(x, \pi)$. In the second subsection, we proceed to determine the global optimal solution $I^*(x, \pi^*)$ by finding the optimal $\pi^*$. In the third subsection, we provide an in-depth discussion based on the solution under an explicit utility function.

3.1. **The optimal health insurance with a fixed premium.** We keep the insurance premium $\pi$ fixed and consider the following optimal health insurance problem:
Problem $P_2$.  

\[
\begin{aligned}
\max_I & \quad g(I(X), \pi) \\
\text{s.t.} & \quad \pi \geq (1 + \rho)E[I(X)], \\
& \quad 0 \leq I(X) \leq \min(K, X).
\end{aligned}
\] (3.4)

To make our problem solvable and manageable, we give the following assumptions first.

**Assumption 1.** As in Eq. (2.1), the utility function has the following form:

\[
U(h, w, y) = h \cdot U(w, y) = h \cdot (\delta V(w) + U(y)),
\]

and $U(y)$ and $V(w)$ satisfy $U' > 0$, $U'' < 0$, $V' > 0$, and $V'' < 0$. Here, $U'$ and $U''$ denote the first and second order derivatives of $U(\cdot)$ respectively. The definitions of $V'$ and $V''$ are similar.

**Assumption 2.** The function $V(w)$ defined in in Eq. (2.1) satisfies the following inequality

\[
\begin{aligned}
& w - (V')^{-1}(\lambda (1 + \rho) p_1 f(x|\Theta = \theta_i) + \lambda (1 + \rho) \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} f(x|\Theta = \theta_i) \delta + \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} f(x|\Theta = \theta_i) \delta) \\
& \quad \geq 0,
\end{aligned}
\] (3.5)

where $(V')^{-1}(\cdot)$ is the inverse function of $V'(\cdot)$.

We point out that Assumption 2 guarantees that the solution to optimization problem $P_2$ exists.

With the help of the Lagrange multiplier method and optimal control technique, we can establish and prove the following proposition:

**Proposition 1.** Under Assumption 1 and Assumption 2, the solution to the optimization problem $P_2$ is

\[
I^*(x) = \begin{cases} 
0, & x \leq d \\
-x - d, & d < x \leq d + K \\
K, & x > d + K
\end{cases}
\] (3.6)

where $d$ is a nonnegative deductible, and satisfies

\[
(1 + \rho)E[I^*(X)] = (1 + \rho) \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} \int_0^\infty I^*(x) dF(x|\Theta = \theta_i) \\
+ (1 + \rho) p_1 \int_0^\infty I^*(x) dF(x|\Theta = 0) \\
= \pi.
\] (3.7)

**Proof.** $I(x)$ is written by $I$ for notational simplicity. Let

\[
L = p_1 \int_0^\infty U(h, w - x + I(x), y - \pi) dF(x|\Theta = 0) \\
+ \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} \int_0^\infty U(h - \theta_i, w - x + I(x), \alpha y - \pi) dF(x|\Theta = \theta_i),
\]

\[
- \lambda \left( (1 + \rho) p_1 \int_0^\infty IDF(x|\Theta = 0) + (1 + \rho) \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} \int_0^\infty IDF(x|\Theta = \theta_i) - \pi \right),
\] (3.8)
where \( \lambda \) is the Lagrange multiplier. The Hamiltonian function corresponding to the Problem \( \mathbf{P}_2 \) is
\[
H = p_1 U(h, w - x + I(x), y - \pi) f(x|\Theta = 0) \\
+ \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} U(h - \theta_i, w - x + I(x), \alpha y - \pi) f(x|\Theta = \theta_i) \\
- \lambda (1 + \rho) p_1 I f(x|\Theta = 0) - \lambda (1 + \rho) \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} I f(x|\Theta = \theta_i),
\]
(3.9)
The Hamiltonian function is concave in \( I \). From the first-order condition \( \frac{\partial H}{\partial I} = 0 \), we have
\[
\begin{align*}
p_1 U_2'(h, w - x + I(x), y - \pi) f(x|\Theta = 0) \\
+ \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} U_2'(h - \theta_i, w - x + I(x), \alpha y - \pi) f(x|\Theta = \theta_i) \\
= \lambda (1 + \rho) p_1 I f(x|\Theta = 0) + \lambda (1 + \rho) \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} f(x|\Theta = \theta_i),
\end{align*}
\]
(3.10)
where \( U_2'(\cdot, \cdot, \cdot) \) denotes partial derivative such as \( U_2' = \frac{\partial U(h, w, y)}{\partial w} \times \frac{\partial w}{\partial I} \). Solving (3.10), we obtain the optimal solution \( I^* \) to Problem \( \mathbf{P}_2 \) under Assumption 1.

\[
\begin{align*}
p_1 I f(x|\Theta = 0) & \delta V'(w - x + I(x)) \\
+ \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} I f(x|\Theta = \theta_i) & \delta V'(w - x + I(x)) \\
= \lambda (1 + \rho) p_1 I f(x|\Theta = 0) + \lambda (1 + \rho) \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} f(x|\Theta = \theta_i).
\end{align*}
\]
(3.11)
Solving (3.11) for \( I^* \), we have
\[
I^* = (x - d)^+ \\
d = w - (V')^{-1} \left( \frac{\lambda (1 + \rho) p_1 I f(x|\Theta = 0) + \lambda (1 + \rho) \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} f(x|\Theta = \theta_i)}{p_1 I f(x|\Theta = 0) \delta + \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} f(x|\Theta = \theta_i) \delta} \right).
\]
(3.12)
where \( (V')^{-1}(\cdot) \) is the inverse function of \( V'(\cdot) \), and it surely exists since \( V'(\cdot) \) is strictly decreasing. Under Assumption 2, \( d \) is a nonnegative deductible. It also satisfies
\[
(1 + \rho) E[I^*(X)] = (1 + \rho) \sum_{\theta_i \in [\theta_l, \theta_u]} p_{\theta_i} \int_0^\infty I^*(x) dF(x|\Theta = \theta_i) \\
= (1 + \rho) p_1 \int_0^\infty I^*(x) dF(x|\Theta = 0) \\
= \pi.
\]
Under the constraint \( 0 \leq I(X) \leq \min(K, X) \) in Problem \( \mathbf{P}_2 \), we obtain the optimal solution as follows:
\[
I^* = \begin{cases} 
0, & x \leq d \\
\ x - d, & d < x \leq \min(K, x) \\
K, & x > \min(K, x)
\end{cases}
\]
which can be rewritten as
\[
I^* = \begin{cases} 
0, & x \leq d \\
 x - d, & d < x \leq d + K \\
K, & x > d + K
\end{cases}
\]

Thus we have proved the proposition. \(\square\)

3.2. Determination of the optimal health insurance. In the previous subsection, the insurance premium is assumed to be fixed, so \(d\) can be considered as a function of \(\pi\). In this subsection, we proceed to determine the global optimal health insurance. Once we obtain the optimal deductible \(d^*\), we eventually derive the solution to Problem \(P_1\).

Proposition 2. Under Assumption 1 and Assumption 2, let \(I^*(X)\) be the optimal solution to Problem \(P_2\) as defined by (3.6), and suppose that \(I^*(X)\) satisfies the following Eq. (3.13), then it also solves Problem \(P_1\). Meanwhile, the optimal premium \(\pi^*\) is \((1 + \rho)E[I^*(X)]\).

\[
0 = (1 - p_1 - p_2)hU'(y - (1 + \rho)E[I^*(x)])Q \\
+ p_1 \int_0^\infty h\delta V'(w - x + I^*(x)) \times (-1_{\{d < x \leq d + K\}}) \\
+ hU'(y - (1 + \rho)E[I^*(x)])QdF(x|\Theta = 0) \\
+ \sum_{\theta_i \in \theta_1, \theta_2} p_{\theta_i} \int_0^\infty (h - \theta_i)\delta V'(w - x + I^*(x)) \times (-1_{\{d < x \leq d + K\}}) \\
+ (h - \theta_i)U'(\alpha y - (1 + \rho)E[I^*(x)])QdF(x|\Theta = \theta_i) \tag{3.13}
\]

where
\[
Q = (1 + \rho)\left[p_1(F(d + K|\Theta = 0) - F(d|\Theta = 0)) + \sum_{\theta_i \in \theta_1, \theta_2} p_{\theta_i}(F(d + K|\Theta = \theta_i) - F(d|\Theta = \theta_i))\right].
\]

Proof. \(I(x)\) is written by \(I\) for notational simplicity. From Proposition 1, we have \(\pi^* = (1 + \rho)E[I^*]\). Let
\[
N = E[U(X, \Theta, I^*, (1 + \rho)E[I^*])] \\
= (1 - p_1 - p_2)U(h, w, y - (1 + \rho)E[I^*]) \\
+ p_1 \int_0^\infty U(h, w - x + I^*, y - (1 + \rho)E[I^*])dF(x|\Theta = 0) \\
+ \sum_{\theta_i \in \theta_1, \theta_2} p_{\theta_i} \int_0^\infty U(h - \theta_i, w - x + I^*, \alpha y - (1 + \rho)E[I^*])dF(x|\Theta = \theta_i).
\]

\(U(h, w, y) = h \cdot U(w, y) = h \cdot (\delta V(w) + U(y))\)

From the first order condition \(\frac{\partial N}{\partial d} = 0\), we have Eq. (3.13). Thus we prove the proposition. \(\square\)

Remark 1. Actually, we can calculate the optimal deductible \(d^*\) by solving Eq. (3.13). Therefore, the global optimal health insurance is determined accordingly.
3.3. Discussion on Optimal Solution under Explicit Utility Function. In the preceding, we have solved the optimal health insurance problem under the general utility framework $U(h, w, y) = h \cdot (\delta V(w) + U(y))$. In this subsection, we provide an in-depth discussion based on the solution under the explicit utility function $U(h, w, y) = h \cdot (\log(y) + \delta \log(w))$. As in Viscusi and Evans [34] and Levy and Nir [20], the log used in our paper represents the natural logarithm.

**Proposition 3.** Under Assumption 2 and suppose that the utility function $U(h, w, y)$ has the form $U(h, w, y) = h \cdot (\log(y) + \delta \log(w))$. Then Eq. (3.13) can be rewritten as follows:

$$
\frac{p_1 h \delta}{w - d} [F(d + K|\Theta = 0) - F(d|\Theta = 0)] + \frac{\delta}{w - d} \sum_{\theta_i \in \Theta} p_{\theta_i} (h - \theta_i) [F(d + K|\Theta = \theta_i) - F(d|\Theta = \theta_i)] \\
= \frac{1}{y - (1 + \rho)E[I^*]} \left[ (1 - p_1 - p_2) h Q + \frac{p_1 h Q}{y - (1 + \rho)E[I^*]} + \frac{Q}{\alpha y - (1 + \rho)E[I^*]} \sum_{\theta_i \in \Theta} p_{\theta_i} (h - \theta_i) \right].
$$

(3.14)

**Proof.**

RHS of (3.13)

$$
= (1 - p_1 - p_2) h U'(y - (1 + \rho)E[I^*(x)]) Q \\
- p_1 \int_{d}^{d+K} h \delta V'(w-d) dF(x|\Theta = 0) \\
+ p_1 \int_{0}^{\infty} h U'(y - (1 + \rho)E[I^*(x)]) Q dF(x|\Theta = 0) \\
- \sum_{\theta_i \in \Theta} \int_{d}^{d+K} (h - \theta_i) \delta V'(w-d) dF(x|\Theta = \theta_i) \\
+ \sum_{\theta_i \in \Theta} \int_{0}^{\infty} (h - \theta_i) U'(\alpha y - (1 + \rho)E[I^*(x)]) Q dF(x|\Theta = \theta_i),
$$

$$
= (1 - p_1 - p_2) h Q \\
\frac{p_1 h \delta}{w - d} [F(d + K|\Theta = 0) - F(d|\Theta = 0)] \\
+ \frac{p_1 h Q}{y - (1 + \rho)E[I^*]} \\
- \frac{\delta}{w - d} \sum_{\theta_i \in \Theta} p_{\theta_i} (h - \theta_i) [F(d + K|\Theta = \theta_i) - F(d|\Theta = \theta_i)] \\
+ \frac{Q}{\alpha y - (1 + \rho)E[I^*]} \sum_{\theta_i \in \Theta} p_{\theta_i} (h - \theta_i) \\
= 0,
$$

which can be rearranged into Eq. (3.14). □

**Remark 2.** The three terms on the right hand side of Eq. (3.14) represent the marginal utilities respectively under case (i), (ii) and (iii).
Eq. (3.14) can be interpreted economically. The right-hand side (RHS) is the marginal utility cost of paying additional premium, and the left-hand side (LHS) is the marginal utility benefit of receiving additional indemnity. Note that Young [37], Zhou and Wu [42] provided similar analysis. Compare our results with Zhou and Wu [42], the significant difference lies in the composition of the RHS, which can be decomposed into three parts, i.e. (1) marginal utility under case (i); (2) marginal utility under case (ii); and (3) marginal utility under case (iii). Compare our results with most of existing literature on optimal insurance, our results distinguish wealth from income. In other words, the short-term health insurance premium is paid from income, monthly or quarterly. When shocks to insured’s health status occur, especially in the case that the insured gets severe or critical illness, his/her wealth would be decreased due to medical expenditure. Therefore, the marginal utility cost of paying additional premium depends on \( y \) (RHS) while the marginal utility benefit of receiving additional indemnity depends on \( w \) (LHS).

4. Numerical Analysis. In this section, we use numerical examples to illustrate the optimal health insurance contract and the optimal premium under specific loss distributions. Then, we examine the impact of various key factors on the optimal indemnity design and the optimal premium.

4.1. Simplified Model for Illustration. Our paper establishes a framework under individual level trade-off among health, wealth and income. To reflect the health insurance contract in reality, we place an upper limit on coverage. For illustration purpose, we consider three basic cases that would occur during the term of health insurance in numerical analysis, i.e. (i) the insured does not get any disease with probability \( 1 - p_1 - p_2 \); (ii) the insured gets mild illness and recovers by brief therapy with probability \( p_1 \); (iii) the insured gets severe or critical illness and his/her health status degenerates by \( \Theta = \theta \) with probability \( p_2 \). Under the explicit utility \( U(h, w, y) = h \cdot (\log(y) + \delta \log(w)) \), we obtain the optimal health insurance as given by Eq. (3.6) and the optimal deductible is determined by

\[
\begin{align*}
& \frac{p_1 h \delta}{w - d} [F(d + K|\Theta = 0) - F(d|\Theta = 0)] \\
& + \frac{\delta}{w - d} p_2 (h - \theta) [F(d + K|\Theta = \theta) - F(d|\Theta = \theta)] \\
& = \frac{(1 - p_1 - p_2) h J}{y - (1 + \rho) E[I^*]} + \frac{p_1 h J}{y - (1 + \rho) E[I^*]} + \frac{\mathcal{J}}{\alpha y - (1 + \rho) E[I^*]} p_2 (h - \theta) \\
& = \frac{(1 - p_2) h J}{y - (1 + \rho) E[I^*]} + \frac{\mathcal{J}}{\alpha y - (1 + \rho) E[I^*]} p_2 (h - \theta), \tag{4.15}
\end{align*}
\]

where

\[
\mathcal{J} = (1 + \rho) \left[ p_1 (F(d + K|\Theta = 0) - F(d|\Theta = 0)) + p_2 (F(d + K|\Theta = \theta) - F(d|\Theta = \theta)) \right].
\]

Remark 3. Eq. (4.15) is actually a revised version of Eq. (3.14) under the illustration model set-up.

4.2. Parameter Values. In an effort to fulfill China’s commitment to the United Nations 2030 Agenda for Sustainable Development, the Political Bureau of the Chinese Community Party adopted the “Healthy China 2030” plan as part of a comprehensive strategy to improve the general health of Chinese people (Wu et
The main goals include (i) improving the health care system and service; (ii) advocating health lifestyle; (iii) developing health industry; (iv) enhancing environmental protection and green growth; and (v) establishing health-related technological and structural innovation.

Under this strategy, it is imperative to carry out large-scale statistical survey and effectively measure the health level of residents. Zhang [39] investigated the health level of the elderly in China based on the China Health and Retirement Longitudinal Study (CHARLS) data. We use the research results of existing literatures, such as the mean score of health status (Zhang, [39]) and health life expectancy (Qiao and Hu, [29]) in our numerical analysis.

Moreover, the China Association of Actuaries released the first “China Life Insurance Critical Illness Morbidity Table” in May 2013, and revised it in May 2020. The formal 2020 version, including the national morbidity table and Guangdong-Hong Kong-Macao Greater Bay Area morbidity table, is scheduled to be in force soon. Our paper takes reference of the illness incidence data in the morbidity table.

In addition, people’s attention on health has stimulated the demand for health insurance products, which has experienced rapid growth from 2010 to 2019 in China. According to the industry statistical data provided by China Banking and Insurance Regulatory Commission, health insurance premium incomes surged from 67.7 billion yuan to 706.6 billion yuan in the past decade. Seriously affected by the outbreak of COVID-19 around the world, health insurance premium income soars in the first quarter of 2020 in China, by 43.97% on a year-on-year basis. Over 70% of the customers are married middle-aged and elderly residents. Therefore, we focus on these people in numerical illustration.

Then, we take reference to the Chinese Family Panel Studies (CFPS) data, Chinese Household Income Projects (CHIP) data, and the numerical illustration in Bernard et al. [4] to decide the values of $\rho$, $K$, $w$, $y$, $\delta$, and $\alpha$. Throughout the numerical analysis, unless otherwise stated, the basic parameters are given in Table 1.

### Table 1. Parameter values for the illustration model

| Parameter                                                   | Symbol | Value |
|-------------------------------------------------------------|--------|-------|
| the probability that the insured gets mild illness and $\Theta = 0$ | $p_1$  | 0.50  |
| the probability that the insured gets severe illness or critical illness and $\Theta = \theta$ | $p_2$  | 0.05  |
| safety loading                                              | $\rho$ | 0.3   |
| upper limit on coverage                                     | $K$    | 10    |
| initial wealth                                              | $w$    | 15    |
| initial income                                              | $y$    | 4     |
| discount factor                                             | $\delta$ | 0.9 |
| reduced income level                                        | $\alpha$ | 0.5 |
| initial health status                                       | $h$    | 0.9   |
| severity of health degeneration after treatment             | $\theta$ | 0.2 |

4.3. Example 1: Exponential Loss Distribution.

**The optimal indemnity under the Exponential loss distribution**

We assume that the loss $X$ follows the exponential distribution conditional on occurrence of shocks to his/her health status, and the density function is

$$f(x|\Theta) = me^{-mx},$$
where the intensity parameter \( m \) varies w.r.t. \( \Theta \). Thus, we have
\[
f(x|\Theta = 0) = m_0 e^{-m_0 x},
\]
and
\[
f(x|\Theta = \theta) = m_1 e^{-m_1 x}.
\]
Let \( m_0 = 0.2 \) and \( m_1 = 0.1 \), we have \( d = 2.33 \), and
\[
I^*(x) = \begin{cases} 
0, & x \leq 2.33 \\
2.33 - x, & 2.33 < x \leq 12.33 \\
10, & x > 12.33 
\end{cases}
\]
which is shown in Figure 1.

When the medical cost is less than or equal to 2.33, the insured cannot get any reimbursement from the insurer. While, the insured can claim his/her medical spending beyond 2.33 when it varies from 2.33 to 12.33. The maximum amount of reimbursement is capped at 10 due to the pre-specified upper limit on coverage, which distinguishes our results from the solution to Arrow’s model in Arrow [1].

Sensitivity test: Exponential loss distribution

We perform sensitivity analysis under Exponential loss distribution, with respect to income \( y \), safety loading \( \rho \), degree of health degeneration \( \theta \), the upper limit on coverage \( K \), health status \( h \), and wealth \( w \), respectively. Figure 3 to Figure 8 depict the results.

From Figure 3, we observe that the optimal deductible \( d^* \) decreases as \( y \) increases, and it reduces to zero when \( y \) takes value 6.16 which is the so-called “critical value”. This result indicates that: (i) it is optimal for the insured to purchase full coverage health insurance when his/her income is greater than or equal to the critical value; (ii) it is optimal for the insured to purchase partial coverage health insurance when his/her income is less than or equal to the critical value. This conclusion agrees with the finding in Lee [19].

From Figure 4, we also observe that a greater \( \rho \) leads to a higher \( d^* \). When \( \rho \) varies from 0 to 1, \( d^* \) always takes positive value. This result means that there is no sudden switch from full to partial coverage insurance due to the slight increase of the safety loading factor from zero.

Figure 5 shows that the degree of health degeneration \( \theta \) does not have obvious influence on \( d^* \) in comparison with the influence of the income \( y \) or safety loading \( \rho \) on \( d^* \). However, a greater \( \theta \) does lead to a higher \( d^* \). This shows that the uncertainty about diagnosis, prognosis, and the impact of treatment clearly influences the optimal indemnity.

Figure 6 shows that the optimal deductible \( d^* \) increases with respect to \( K \). Especially, there is a critical value of \( K \). When \( K \) takes value 4.25, \( d^* \) is equal to zero. This result means that: (i) it is optimal for the insured to purchase full coverage health insurance when the upper limit on coverage is lower than or equal to the critical value; (ii) it is optimal for the insured to purchase partial coverage health insurance when the upper limit on coverage is greater than or equal to the critical value.

Figures 7 and 8 show that the current health status \( h \) and wealth \( w \) both have relatively small impact on the optimal deductible. We observe that the optimal deductible \( d^* \) decreases with respect to \( h \), while it increases with respect to \( w \).
4.4. Example 2: Pareto Loss Distribution.

The optimal indemnity under the Pareto loss distribution

We assume the loss $X$ follows the Pareto distribution conditional on occurrence of shocks to his/her health status, and the density function is

$$f(x|\Theta) = \frac{\beta \xi^\beta}{(\xi + x)^{\beta+1}},$$

where the intensity parameter $m$ varies w.r.t. $\Theta$. Hence, we have

$$f(x|\Theta = 0) = \frac{\beta_0 \xi_0^\beta}{(\xi_0 + x)^{\beta_0+1}},$$

and

$$f(x|\Theta = \theta) = \frac{\beta_1 \xi_1^\beta}{(\xi_1 + x)^{\beta_1+1}}.$$  

Let $\beta_0 = \beta_1 = 3$, $\xi_0 = 10$, $\xi_1 = 20$, we have $d = 1.42$, and

$$I^*(x) = \begin{cases} 
0, & x \leq 1.42 \\
 x - 1.42, & 1.42 < x \leq 11.42 \\
 10, & x > 11.42
\end{cases}$$  \hspace{1cm} (4.17)$$

which is shown in Figure 2.

Similarly to the $I^*(X)$ under the Exponential loss distribution, the maximum reimbursement herein is also capped at 10 because of the risk constraint $I(X) \leq K$. More specifically, the insured pays out-of-pocket cost 1.42, and the insurer pays for the medical expenditure beyond 1.42 up to 11.42.

The exponential distribution and Pareto distribution we use in the numerical example have same mean but different variance. Due to the different level of dispersion and other probabilistic properties, the optimal deductible under the Pareto distribution is much lower than that under the exponential distribution.

Sensitivity test: Pareto loss distribution

We perform sensitivity analysis under the Pareto loss distribution, with respect to income $y$, safety loading $\rho$, degree of health degeneration $\theta$, the upper limit on coverage $K$, health status $h$, and wealth $w$, respectively. The results are shown in Figures 9 to 14.

We observe that the effects of the key factors on $d^*$ are similar to these under the Exponential loss distribution. When these factors vary within the same range, $d^*$ usually takes smaller value. Moreover, the critical values for $y$ and $K$ are 5.36 and 5.21, respectively. Similarly, both the current health status $h$ and the wealth $w$ have less significant impact on the optimal deductible.

5. Conclusion. In this paper we investigate the optimal health insurance problem with constraints under an explicit utility of health, wealth and income. The preference framework we use herein gives a better description of the trade-off among health, wealth and income in reality. We derive and show the optimal indemnity design above a deductible up to a cap. Furthermore, the optimal deductible is shown to depend on the health status, wealth and income. With consideration of illness incidence, medical expenditure and health level of residents, we illustrate the optimal health insurance contract by numerical examples. In addition, we examine the impact of the key parameters, including exogenous shocks to health, safety
loading and income, on the optimal indemnity design. Our research results give new insights to insurance contract design for health insurance industry.
Figure 3. Sensitivity of $d^*$ w.r.t. income $y$ under Exponential loss distribution.

Figure 4. Sensitivity of $d^*$ w.r.t. safety loading $\rho$ under Exponential loss distribution.
Figure 5. Sensitivity of $d^*$ w.r.t. degree of health degeneration $\theta$ under Exponential loss distribution.

Figure 6. Sensitivity of $d^*$ w.r.t. upper limit on coverage $K$ under Exponential loss distribution.
Figure 7. Sensitivity of $d^*$ w.r.t. current health status $h$ under Exponential loss distribution.

Figure 8. Sensitivity of $d^*$ w.r.t. current wealth $w$ under Exponential loss distribution.
Figure 9. Sensitivity of $d^*$ w.r.t. income $y$ under Pareto loss distribution.

Figure 10. Sensitivity of $d^*$ w.r.t. safety loading $\rho$ under Pareto loss distribution.
Figure 11. Sensitivity of $d^*$ w.r.t. degree of health degeneration $\theta$ under Pareto loss distribution.

Figure 12. Sensitivity of $d^*$ w.r.t. upper limit on coverage $K$ under Pareto loss distribution.
Figure 13. Sensitivity of $d^*$ w.r.t. current health status $h$ under Pareto loss distribution.

Figure 14. Sensitivity of $d^*$ w.r.t. current wealth $w$ under Pareto loss distribution.
REFERENCES

[1] K. J. Arrow, Uncertainty and the welfare economics of medical care, *American Economic Review*, 53 (1963), 941–973.
[2] K. J. Arrow, *Essays in the Theory of Risk-Bearing*, North-Holland Publishing Co., Amsterdam-London, 1970.
[3] K. J. Arrow, Optimal insurance and generalized deductibles, *Scandinavian Actuarial Journal*, 1974 (1974), 1–42.
[4] C. Bernard, X. He, J.-A. Yan and X. Y. Zhou, Optimal insurance design under rank-dependent expected utility, *Mathematical Finance*, 25 (2015), 154–186.
[5] T. J. Besley, Optimal reimbursement health insurance and the theory of Ramsey taxation, *Journal of Health Economics*, 7 (1988), 321–336.
[6] S. Bhargava, G. Loewenstein and J. Sydnor, Choose to lose: Health plan choices form a menu with dominated option, *Quarterly Journal of Economics*, 132 (2017), 1319–1372.
[7] A. Blomqvist, Optimal non-linear health insurance, *Journal of Health Economics*, 16 (1997), 303–321.
[8] N. A. Doherty and H. Schlesinger, Optimal insurance in incomplete markets, *Journal of Political Economy*, 91 (1983), 1045–1054.
[9] D. Doiron and N. Kettlewell, Family formation and demand for health insurance, *Health Economics*, 29 (2020), 523–533.
[10] R. P. Ellis, S. Jiang and W. G. Manning, Optimal health insurance for multiple goods and time periods, *Journal of Health Economics*, 41 (2015), 89–106.
[11] Y. Fan, The solution of the optimal insurance problem with background risk, *Journal of Function Spaces*, 2019, Art. ID 2759398, 5 pp.
[12] A. Finkelstein, E. F. P. Luttmer and M. J. Notowidigdo, What good is wealth without health? The effect of health on the marginal utility of consumption, *Journal of the European Economic Association*, 11 (2013), 221–258.
[13] A. Finkelstein, N. Hendren and M. Shepard, Subsidizing health insurance for low-income adults: evidence from Massachusetts, *American Economic Review*, 109 (2019), 1530–1567.
[14] V. R. Fuchs and R. J. Zeckhauser, Valuing health: A priceless commodity, *American Economic Review Papers and Proceedings*, 77 (1987), 263–268.
[15] M. Gerfin, Health insurance and the demand for healthcare, *Oxford Research Encyclopedia of Economics and Finance*, 1 (2019), 1–22.
[16] C. Gollier and H. Schlesinger, Arrow’s theorem on the optimality of deductibles: A stochastic dominance approach, *Economic Theory*, 7 (1996), 359–365.
[17] R. E. Hall and C. I. Jones, The value of life and the rise in health spending, *Quarterly Journal of Economics*, 122 (2007), 39–72.
[18] T. Y. Ho, P. A. Fishman and Z. B. Zabinsky, Using a game-theoretic approach to design optimal health insurance for chronic disease, *IISE Transactions on Healthcare System Engineering*, 9 (2019), 26–40.
[19] K. Lee, Wealth, income, and optimal insurance, *Journal of Risk and Insurance*, 74 (2007), 175–184.
[20] M. Levy and A. R. Nir, The utility of health and wealth, *Journal of Health Economics*, 31 (2012), 379–392.
[21] T. Li and X. Wang, Connotation and realizing route of Healthy China, *Health Economics Research*, 1 (2016). 4–10.
[22] Z. Lu, S. Meng, L. Liu and Z. Han, Optimal insurance design under background risk with dependence, *Insurance: Mathematics and Economics*, 80 (2018), 15–28.
[23] H. Markowitz, The utility of wealth, *Journal of Political Economy*, 60 (1952), 151–156.
[24] K. S. Moore and V. R. Young, Optimal insurance in a continuous-time model, *Insurance: Mathematics and Economics*, 39 (2006), 47–68.
[25] J. Mossin, Aspects of rational insurance purchasing, *Journal of Political Economy*, 76 (1968), 552–568.
[26] C. E. Phelps, *The Demand for Health Insurance: A Theoretical and Empirical Investigation*, Rand Corporation, Santa Monica, Report R-1054-OEO, 1973.
[27] J. S. Pliskin, D. S. Shepard and M. C. Weinstein, Utility functions for life years and health status, *Operations Research*, 28 (1980), 206–224.
[28] S. D. Promislow and V. R. Young, Unifying framework for optimal insurance, *Insurance: Mathematics and Economics*, 36 (2005), 347–364.
[29] X. Qiao and Y. Hu, Health expectancy of Chinese elderly and provincial variabilities, *Population & Development*, 23 (2017), 2–18.
[30] A. Raviv, The design of an optimal insurance policy, *American Economic Review*, 69 (1979), 84–96.
[31] V. L. Smith, Optimal insurance coverage, *Journal of Political Economy*, 76 (1968), 68–77.
[32] M. Spence and R. Zeckhauser, Insurance, information, and individual action, *American Economic Review*, 61 (1971), 380–387.
[33] S. Wu, Y. Luo, X. Qiu and M. Bao, Building a healthy China by enhancing physical activity: Priorities, challenges, and strategies, *Journal of Sport and Health Science*, 6 (2017), 125–126.
[34] W. K. Viscusi and W. N. Evans, Utility functions that depend on health status: Estimates and economic implications, *American Economic Review*, 80 (1990), 353–374.
[35] Z. Q. Xu, X. Y. Zhou and S. C. Zhuang, Optimal insurance under rank-dependent utility and incentive compatibility, *Mathematical Finance*, 29 (2019), 659–692.
[36] J. Yan, Why does China’s government basic medical insurance system enhance the development of commercial health insurance?, *Chinese Review of Financial Studies*, 2 (2018), 1-15+221.
[37] V. R. Young, Optimal insurance under Wang’s premium principle, *Insurance: Mathematics and Economics*, 25 (1999), 109–122.
[38] R. Zeckhauser, Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives, *Journal of Economic Theory*, 2 (1970), 10–26.
[39] Y. Zhang, Dynamic research on the health level of the elderly in China under the background of Healthy China: Based on CHARLS data empirical analysis, *Northwest Population*, 40 (2018), 50–59.
[40] Y. Zhang and Y. Wu, Optimal health insurance and trade-off between health and wealth, *Journal of Applied Mathematics*, 2020, Art. ID 2658213, 9 pp.
[41] Y. Zheng, T. Vukina and X. Zheng, Estimating asymmetric information effects in health care with uninsurable costs, *International Journal of Health Economics and Management*, 19 (2019), 79–98.
[42] C. Zhou and C. Wu, Optimal insurance under the insurer’s risk constraint, *Insurance: Mathematics and Economics*, 42 (2008), 992–999.
[43] C. Zhou, W. Wu and C. Wu, Optimal insurance in the presence of insurer’s loss limit, *Insurance: Mathematics and Economics*, 46 (2010), 300–307.

Received July 2020; revised October 2020.

E-mail address: yan.zhang2017@gdufs.edu.cn
E-mail address: y.wu@curtin.edu.au
E-mail address: yaohaixiang@gdufs.edu.cn