Journal of Regional Economics 2023 2 (1) 53-68

Journal of Regional Economics

Homepage: https://www.anserpress.org/journal/jre

Portfolio Allocation with Medical Expenditure Risk-A Life Cycle Model and Machine Learning Analysis

You Du [a], Weige Huang [b,*]

[a] *Department of Economics, Finance, and Accounting, Robbins College of Business and Entrepreneurship, Fort Hays State University, Hays, KS, USA*
[b] *Faculty of Economics, Shenzhen MSU-BIT University, Shenzhen, China*

ABSTRACT

This paper explores how the medical expenditure risk affects the households' portfolio choice across health status theoretically in a life cycle model and empirically using machine learning methods. Medical expenditure risk, as a background risk, has the potential to influence households' financial decisions. A higher medical expenditure risk leads to a larger fluctuation and more uncertainty in households' consumption and therefore utility. As a result, risk-free assets become more attractive. Our machine learning analysis provides evidence that aligns with the predictions of the theoretical life cycle model. Specifically, households with better health hold a larger proportion of stocks in their portfolios. Furthermore, when facing increased medical expenditure risk, households in good health demonstrate a greater willingness to invest in safe assets.

KEYWORDS

Household finance; medical expenditure risk; life cycle model; machine learning

* Corresponding author: Weige Huang
E-mail address: huangwg@smbu.edu.cn

ISSN 2972-3434
doi: 10.58567/jre02010005

Received 25 May 2023; Accepted 6 July 2023; Available online 11 October 2023

1. Introduction

This paper investigates the effect of medical expenditure risk on households' financial allocations across health in the frame of a life cycle model and machine learning analysis. Withthe increasing concern over high healthcare costs, especially out-of-pocket medical expenses, it naturally becomes a significant factor influencing individuals' financial decision-making. Medical expenditure risk, which cannot be fully insured in the market, can be seen as a background risk for households. However, existing literature primarily focuses on the level or value of medical expenses rather than the associated risk (e.g., Rosen and Wu (2004), Ryan Edwards (2008), and Ayyagari and He (2017)). While the level of medical expenditure affects household budgets and their ability to invest in financial assets, it alone cannot explain the allocation of risky shares,i.e., the proportion of risky financial investments relative to total financial assets. In contrast, out-of-pocket medical expenditure risk (hereafter OOP risk), which refers to the fluctuation of the out-of-pocket medical expenditure over life profiles, is able to reallocate households' financial assets structure. The mechanism of health expenditure risk is as follows: A higher medical expenditure risk indicates greater fluctuations and increased uncertainty in households' consumption and utility. Consequently, risk-free assets become more appealing as they provide a hedge againstmedical expenditure risk, ensuring smoother expected consumption and utility for households over time.

In this paper, our empirical findings are based on the RAND Health and Retirement Study (HRS) data (hereafter HRS). The HRS dataset is a comprehensive and well-organized collection of information, encompassing various variables such as demographics, health, health insurance, Social Security, pensions, family structure, retirement plans, expectations, and employment history. Additionally, the dataset includes imputations for income, assets, and medical expendituresto enhance its usability and accuracy. The analysis of the HRS data reveals interesting patterns.It suggests that households with better health exhibit lower medical expenditure risk throughouttheir lifetime compared to those with poorer health. Furthermore, the relationship between an individual's medical expenditure risk and their portfolio choice appears to be contingent on their health status. This implies that the impact of medical expenditure risk on portfolio allocation varies depending on whether the individual is classified as healthy or not. By leveraging the rich and extensive HRS dataset, this paper provides empirical evidence highlighting the interplay between medical expenditure risk, health status, and households' portfolio choices.

To intensively evaluate the contribution of medical spending risk in explaining the portfolio choice across health, a counterfactual exercise is employed: we eliminate the influence of medical expenditure risk, effectively removing this risk factor from households' decision-making. Unsurprisingly,both groups' risky share increase. However the effects are asymmetric between the two groups: the effect is larger on the poor healthy group: the risky portfolio ratio rises from 76.63% to 82.04%, representing a 7 percentage increase. While for the good health group, the ratio increases from 83.35% to 86.99%, with only a 4 percentage rise. Even without the presence of medical expenditure uncertainty, the risky share of the poor health group remains lower than that of the good health group at every age and on average, with a 6% lower allocation. This discrepancy can be attributedto the strong positive correlation between health status and income. Existing literature suggests that good health is associated with higher income (e.g., Angus Deaton (2003), Maria Jose Prados (2017)). Higher income can be viewed as a form of risk-free assets, implying a relatively higher allocation to safe assets compared to total financial assets. As a result, better health, coupled with higher income, leads to a higher allocation to risky assets in the portfolio.

Our paper makes contributions in the following streams. Firstly, it contributes to the literatureby conducting a time series analysis of out-of-pocket (OOP) medical expenditure risk, which is a relatively rare approach. As far as our knowledge extends, the only existing exception is the study conducted by You Du (2021). Previous literature often focuses on examining the level of medical expenditure, overlooking the aspect of "risk." That can not be solely used to explain the risky share choice. In contrast, our study tracks the life profiles of households' medical expenditure and utilizes the variance of this AR (1) process as a measure of "risk." Our paper uses a different dataset (i.e., HRS) from You Du (2021). The HRS dataset provides comprehensive information on respondents' portfolios, health status, and medical expenditures across 13 surveys. This rich dataset enables us to accurately compute medical expenditure risk using an AR (1) model.

Second, our paper is the first time in the literature, to use the life cycle model and machine learning approach. This contributes to the methodology in macro-health economics and household finance fields. In this paper, we adopt two pioneering methods to explore the main question. Theoretical contributions are made by proposing a life cycle model that incorporates medical expenditure risk, capturing its key features. Life cycle models are commonly used in modern Macroeconomics to analyze household consumption and savings decisions under various forms of risk. By incorporating medical expenditure risk into the model, we extend its applicability to the domain of health-related financial decision-making. Empirically, our paper employs an innovativemachine learning analysis, which offers several advantages over traditional inferential statistical methods like regressions. Machine learning approaches provide the opportunity to extract insights from data without imposing prespecified assumptions about variable relationships or structure. Additionally, they enable the modeling of non-linear relationships in high-dimensional spaces, accommodating complex interactions that are often unknown and challenging to specify in advance. To the best of our knowledge, our paper is the first in the literature to study the relationship between portfolio allocations and medical expenditure risk by combining theoretical modeling and empirical research using a machine learning approach. Our findings demonstrate that the empirical evidence derived from machine learning analysis aligns closely with the predictions of the theoretical models.

Third, the paper employs a machine learning approach to validate the predictions made by the structural model. Although machine learning has been widely used in fields such as portfolio optimization (e.g., Ban, El Karoui, and Lim (2018); Perrin and Roncalli, 2020), medicine expenditure(e.g., Nagarjuna et al. (2022); Kaushik et al. (2020)), and health status prediction (e.g., Qin etal. (2020); Tarekegn et al. (2020)), we could not find any literature on the correlation between the three variables studied. Consequently, the paper attempts to address this gap by utilizing a tree-based model (XGBoost) to predict portfolio choice based on medical expenditure risk, health,and earnings. Additionally, the SHAP method interprets the XGBoost results and demonstrates the interaction effects between health status and medical expenditure risk.[1] The machine learninganalysis provides evidence consistent with the structural model's predictions.

To conclude, this paper explores the effect of medical expenditure risk on households' portfolio choices using a life cycle model and a new machine learning technique. This study is the firstto examine portfolio allocations while incorporating health-related factors, using a combination of theoretical modeling and empirical analysis with machine learning. The results from both the structural model and the empirical analysis are in agreement and offer valuable insights. In particular, households with better health tend to invest more in risky assets. Furthermore, when confronted with a higher medical expenditure risk, households in good health are more likely to invest in safer assets.

The rest of the paper is organized as follows. Section 2 summarizes the related literature. Sections 3 and 4 present our structural model and parameters. The benchmark results and counterfactual experiment are shown in Sections 5 and 6, respectively. Section 7 briefly introduces the machine learning method (XGBoost) and SHAP. Section 8 presents data, machine learning results and some robustness checks. Section 9 concludes.

2. Literature Review

This paper makes a significant contribution to the macro-health theoretical literature by combining various strands of research. While previous studies have examined different aspects related to health expenditure, savings, and portfolio allocations, our paper fills a critical gap by integratingthese dimensions within a comprehensive structural model and utilizing a cutting-edge machine learning approach. One of our key findings is the importance of considering medical expenditure risk in understanding households' financial decisions. We show that households can mitigate the impact of high medical expenditure risk by investing in risk-free assets. These assets provide a stable and predictable return, which helps ensure smoother expected consumption and utility forhouseholds over time. By allocating a portion of their portfolio to risk-free assets, households create a hedge against the potential negative effects of medical expenditure shocks, thereby enhancing their financial security and stability.

Previous studies, such as De Nardi et al. (2010) and Kopecky and Koreshkova (2014), have explored the impact of healthcare expenses on savings behavior among seniors. However, they do not explicitly consider the portfolio allocation decisions of households. Similarly, Pang and Warshawsky (2010) and Koijen et al. (2016) investigate portfolio allocations among different financialassets for retired households, but they do not incorporate endogenous health spending into their analyses. Other studies, such as Halliday et al. (2019) and Prados (2017), have explored the implications of health status and health investment on various economic outcomes. However, they do not specifically address the relationship between health expenditure risk and the allocation ofrisky portfolio shares. Yogo (2016) analyzes retirees' financial choices while considering health riskand housing risk, but the specific relationship between health expenditure risk and risky portfolio shares is not explored in detail. In

[1] In many applications, understanding the reasoning behind a model's prediction can be just as significant as the accuracy of the prediction itself. Unfortunately, complex models, such as ensemble or deep learning models, often provide the highest accuracy for large modern datasets but are difficult for even experts to interpret. This creates a conflict between accuracy and interpretability. As a solution, various techniques have been recently proposed to assist users in understanding the predictions of complex models. However, it is often unclear how these methods are interconnected and when one method may be more suitable than another. To tackle this problem, Lundberg and Lee (2017) introduce SHAP, a comprehensive framework designed to provide interpretations of predictions.

contrast to these previous studies, our paper specifically focuses on the relationship between health expenditure risk and the allocation of risky portfolio shares within a comprehensive structural model. Therefore, we provide a more comprehensive understanding of how households integrate health dynamics and financial considerations in theirdecision-making processes.

Our study builds upon the existing empirical literature that explores the relationship between health and portfolio allocations. Rosen and Wu (2004) provide evidence suggesting that health is a significant factor influencing households' decisions among different financial asset categories using the HRS dataset. Edwards (2008) establishes the role of health in explaining the decline in riskyshare among retirees' portfolios. Inkmann et al. (2011) investigate the determinants of annuity market demand with exogenous health spending. Ayyagari and He (2017) find that Medicare beneficiaries increase their risky share after certain medical prescriptions become covered. While these empirical studies offer valuable insights into the link between health and portfolio allocations, our paper extends this line of research by focusing specifically on the impact of medical expenditure risk on households' risky portfolio share. By considering the uncertainty associated with medical expenses, we provide a more comprehensive understanding of how households adjust their risky asset holdings in response to this specific risk. Our findings contribute to the literature by highlighting the importance of medical expenditure risk as a distinct factor influencing households' portfolio decisions and shedding light on the mechanisms through which health and financial choices are interconnected.

3. Full Model

This section describes the life cycle model of portfolio allocation across health with the medical expenditure risk[2]. The benchmark model can be summarized as follows: a retiree enters the model with some initial endowment of financial assets and health capital. In each period, it receives type-specific flat retirement income, gross return of its financial assets and faces two types of uncertainties: financial market risk and medical expenditure risk. Then the senior household chooses consumption and allocates its financial wealth between risk-free bonds and risky stocks in every period to maximize its life time utility.

3.1. Time

In our model, time is discretized, and each period refers to one year. We consider a representative household that enters the model at the age of 65, representing the retirement age. Followingthe literature and implications from HRS data, the maximum of the life period is 81 yearsold in our model[3].

3.2. Type-specific Flat Retirement Income

When the household enters retirement, it receives a fixed type-specific retirement income $Y_t = Y^j$, based on its health status type j, where $j \in$ {poor health, good health}.

3.3. Type-specific Medical Expenditure and Its Risk

The household faces some uncertainty and volatility of out- of- pocket medical spending overtime, which is called medical expenditure risk in this paper.[4] This OOP risk is associated with the retiree's own health type (poor health or good health). As the empirical facts suggest, healthier households have lower medical expenditure risk. An AR(1) process is established to describe the OOP medical expenditure over time. This process allows us to describe the dynamics of the OOP medical expenditure over time, considering the persistence and volatility of this expenditure.

$$log\ (OOP)_{t,i} = \alpha_i + \beta_i log\ (OOP)_{t-1,i} + \epsilon_{t,i}\ (1)$$

where t represents age and i represents an observation in HRS data. The variance of ϵ_i is the primary focus here, which illustrates individual i's OOP risk. A larger $var(\epsilon_i)$ means a higher OOP risk for the individual i. We first estimate Equation 1 for each observation in the HRS data, then the estimated coefficients and the variance of residuals are averaged by health status, i.e. poor health and good health:

---

2 Throughout this paper, the medical expenditure refers to the out-of-pocket medical expenditure, or OOP.

3 In HRS data, we classify the observations above 81 as one age group.

4 Health insurance is not discussed in this paper. We use out-of-pocket (OOP) expenditure to represent the medical expenditure, instead of the total medical bill.

$$\overline{\alpha}_{poor} = \frac{1}{N}\sum_{i=1}^{N} \alpha_{i,i\in poor} \tag{2}$$

$$\overline{\alpha}_{good} = \frac{1}{M}\sum_{j=1}^{M} \alpha_{j,j\in good} \tag{3}$$

$$\overline{\beta}_{poor} = \frac{1}{N}\sum_{i=1}^{N} \beta_{i,i\in poor} \tag{4}$$

$$\overline{\beta}_{good} = \frac{1}{M}\sum_{j=1}^{M} \beta_{j,j\in good} \tag{5}$$

$$\overline{var(\epsilon)}_{i,i\in good} = \frac{1}{N}\sum_{i=1}^{N} var(\epsilon)_{i,i\in good} \tag{6}$$

$$\overline{var(\epsilon)}_{j,j\in poor} = \frac{1}{M}\sum_{j=1}^{M} var(\epsilon)_{j,j\in poor} \tag{7}$$

where N and M are the sample sizes for the poor and good wealth groups, respectively.

Table 1 below summarizes the estimation results[5]. It illustrates that the good health grouphas a smaller variance of log (OOP) over the life cycle, suggesting its OOP risk is lower comparedto the poor health group.

Table 1. AR (1) Process of Log (OOP) (HRS 1999-2017).

	Good Health	Poor Health
$\bar{\alpha}$	5.71	6.28
$\bar{\beta}$	0.16	0.13
$\overline{var}(\epsilon)$	1.86	2.15
Observations	1207	87

3.4. Budget Constraints

Every period, the household receives income and gross return on its financial assets, and thendecides how much to consume and how much to save between financial assets. For clarity, the state variables of financial wealth are named as “asset in bonds Ab_t” and “asset in stocks As_t”. The control variables of financial wealth that the retiree chooses are “savings in bonds Ab_{t+1}” and“savings in stocks As_{t+1}”. Out-of-pocket medical spending is exogenous and follows a Markov chain process over time. The budget constraint is:

$$C_t + Ab_{t+1} + As_{t+1} + OOP_t = Ab_t Rb_t + As_t Rs_t + Y_t \tag{8}$$

The left-hand side is the sum of consumption C_t in period t, next period $t + 1$’s saving between bonds and stocks, and out- of- pocket medical expenditure in period t. The right- hand side isthe household’s cash on hand including its flat income Y_t in period t and the gross return of its financial assets i.e. $Ab_tRb_t + As_tRs_t$. Rb_t is the gross rate of return for bonds and Rs_t is the gross rate of return for stocks.

Financial assets and consumption are assumed to be non-negative in each period. There is noborrowing in this model.

$$0 \leq Ab_t, As_t, C_t, OOP_t, \forall t \tag{9}$$

[5] In HRS data, health status is defined as five levels: poor, fair, good, very good, and excellent. We classify the poor and fair levels as the poor health group, and the other three levels as the good health group in our paper. Due to the data limitation, it generates unbalanced sample sizes for the two groups. However, the sample sizes for both groups are in a reasonable range and are capable of providing statistical inference.

3.4.1. Return for Financial Assets

We follow Yogo (2016) for the financial assets. The gross return of bonds Rb_t is set to 1.025 and the gross return of stocks is defined as: $r_{st} = \overline{r_s}\epsilon_{s,t}$, where $log(\epsilon_{s,t}) \sim \mathcal{N}(-\sigma_s^2/2, \sigma_s^2)$, $\overline{r_s}$ is 1.065, reflecting the equity premium of 4%.

3.5. Utility Function

The utility function is the adoption of You Du (2021), which "highlights the important value of health in examining optimal portfolio choices." The representative household values both health and consumption in its utility function:

$$U(C_t, H_t) = \frac{((1-\alpha)C_t + \alpha H_t)^{1-\sigma}}{1-\sigma} \tag{10}$$

$\alpha \in (0, 1)$ is the health weight in utility. $(1-\alpha)$ is the utility weight on consumption. $\sigma > 1$ is the coefficient of risk averse.

3.6. Household's Problem

Formally, at age t, a type j household is characterized by four state variables[6]: asset in bonds Ab_t, asset in stocks As_t, gross rate of return for stocks Rs_t, and out-of-pocket medical spending OOP_t. The household's problem could be written recursively as:

$$U(C_t, H_t) = \frac{((1-\alpha)C_t + \alpha H_t)^{1-\sigma}}{1-\sigma} \tag{11}$$

subject to

$$C_t = Ab_t Rb_t + As_t Rs_t + Y_t - Ab_{t+1} - As_{t+1} - OOP_t \tag{12}$$

$$OOP_t = \gamma_1 + \gamma_2 OOP_{t-1} + e_t \tag{13}$$

$$0 \leq Ab_t, As_t, C_t, OOP_t, \forall t \tag{14}$$

Parameter $\beta \in (0, 1)$ is the subjective discount factor. E is the expectation operator.

In period t, the household of state (Ab_t, As_t, Rs_t, OOP_t) chooses consumption C_t, saving in bonds Ab_{t+1} and savings in stocks As_{t+1} to maximize the sum of two components in Equation 11. The first part is the period utility directly from current consumption and health. The second component is the discounted expected future value function. Equation 12 is the budget constraint. The law of motion of medical expenditure is in Equation 13. As we discussed earlier, out-of-pocket medical expense is type-specific and follows an AR (1) process. Equation 14 presents the non-negative constraints for consumption and financial assets in every period.

4. Calibration

The parameters of the model is discussed in this section.

4.1. Type-Specific OOP Risk

The OOP risks, denoted by e^j, are specific to the health type j (poor health or good health). Toapproximate the OOP risks as a two-state Markov-chain process, the Tauchen method is employed. Based on the estimation results presented in Table 1 and the HRS data, the type-specific OOP risks are calibrated as follows: for the poor health group, two OOP shocks are calibrated to:$e_{l,1}$ = 2.78 and $e_{l,2}$ = 11.65, with the transition matrix [0.65, 0.35; 0.35, 0.65]; The two OOP shocks for good health type are $e_{h,1}$ = 2.65 and $e_{h,2}$ = 10.94, with the transition matrix [0.69, 0.31; 0.31, 0.69].

4.2. Preference

We select the discount coefficient as 0.96, and the risk aversion coefficient as 4, values commonlyused in the literature. We follow You Du (2021) that the quality of life coefficient is 0.4.

[6] For neatness, we omit the superscript j in this section.

4.3. Assets Prices

We follow Yogo (2016) when choosing the annual asset price parameters. The fixed annualgross return of bonds is 1.025. The average gross return of stocks $\overline{r_s}$ is 1.065 and the standard deviation of stocks σ_s is chosen as 0.18.

4.4. Summary of Calibration

This part summarizes the parameters used in the model. Some parameters come from related literature. For the model specific OOP risk parameters, we calibrate them by using the data in Section 3. Table 2 is a summary.

Table 2. Calibration.

Name	Description	Value	Sources
		Preference	
$\beta\sigma\alpha$	subjective discount factor CRRA coefficient quality of life	0.96 4 0.4	common literaturecommon literatureYou Du (2021)
		Type-specific OOP risk	
e_h	Good health OOP risk	[2.65, 10.94]	calibrated with HRS data
eM_h	Good health transition matrix	[0.69, 0.31; 0.31, 0.69]	calibrated with HRS data
e_l	Poor health OOP risk	[2.78, 11.65]	calibrated with HRS data
eM_l	Poor health transition matrix	[0.65,0.35; 0.35, 0.65]	calibrated with HRS data
		Financial assets	
$\bar{R}_b\ \bar{R}_S\ \sigma_s$	fixed gross bond returnaverage stock return s.d. of stock returns	1.025 1.065 0.18	Yogo (2016) Yogo (2016) Yogo (2016)

5. Benchmark Results

This section presents the simulation results and the discussion. By employing backward induction and numerical methods, we solve this life cycle model and obtain the policy functions of consumption, savings in bonds, and savings in stocks. Then we simulate two shocks in the model, i.e. medical expenditure shocks and gross rate of return for stocks. Table 3 reports the benchmarkmodel's performance. Overall, the model performs well in matching the HRS data, particularly in terms of the mean risky shares for each health group. The simulated risky shares are slightly higher than that in the HRS data (83.35% vs. 82.2% for good health, and 76.63% vs. 74.49% for poor health). This discrepancy can be attributed to the exclusion of additional background risks that are not explicitly considered in the benchmark model.

Table 3. Risky Share (%) across Health: Data VS. Model.

Age Group	Data-Good H	Model-Good H	Data-Poor H	Model-Poor H
Age 65-66	82.94	77.34	82.05	64.90
Age 67-68	84.09	82.86	74.23	74.22
Age 69-70	81.88	85.21	67.42	78.28
Age 71-72	82.12	86.31	74.6	80.43
Age 73-74	82.52	86.42	77.77	80.77
Age 75-76	85.81	86.09	66.34	80.82
Age 77-78	79.22	85.36	78.31	80.46
Age 79-80	78.06	83.22	74.14	78.55
Age 81+	81.09	77.38	71.46	71.28
Average	82.2	83.35	74.49	76.63

6. Counterfactual Experiment----No Medical Risk

In this experiment, both health groups face constant medical expenditures without any uncertainty. The results show that the elimination of medical expense risk leads to an increase in risky shares for both health groups across all time periods. For the good health group, it increases from 83.35% to 86.99%, with a 4.37% rise. For the poor health group, it climbs from 76.63% up to82.04%, with a 7.06 % increase. The medical expenditure risk illustrates asymmetric effects onthese two health groups, and has a larger effect on the poor health group.

Our results also highlight the importance of health status in portfolio allocations. In this economy, the only uncertainty comes from the stock return, which is identical for both groups. Meanwhile, the good health group's risky share is higher than that of the poor health group acrossall ages and on average, with a 6% difference (86.99% vs. 82.04%). This finding suggests that health status can be viewed as an additional form of "safe assets" that influences households' portfolio choices, leading them to allocate a higher proportion of their wealth to risky assets.

Table 4. Risky Share (%) across Health: Benchmark Model VS Experiment.

Age Group	Model-Good H	Exp -Good H	Model-Poor H	Exp -Poor H
Age 65-66	77.34	79.62	64.90	72.81
Age 67-68	82.86	85.68	74.22	81.32
Age 69-70	85.21	88.57	78.28	83.56
Age 71-72	86.31	89.84	80.43	85.78
Age 73-74	86.42	90.56	80.77	86.95
Age 75-76	86.09	90.39	80.82	86.19
Age 77-78	85.36	89.48	80.46	84.37
Age 79-80	83.22	87.29	78.55	82.04
Age 81+	77.38	81.483	71.28	74.79
Average	83.35	86.99	76.63	82.04

7. Machine Learning Methods

This section introduces the main machine learning techniques utilized in the study: eXtreme Gradient Boosting (XGBoost) and Shapley Additive exPlanations (SHAP). Specifically, XGBoost is employed to model the relationship between the risky share (the proportion of stock asset amount to the sum of stock and bond asset amount in an individual's portfolio) and three key factors: medical risk, health status, and earnings.

The decision to use XGBoost and SHAP was based on several factors. Firstly, the relationship between the risky share and factors is not well-understood and is likely to be complex and non-linear. A simple linear regression model would not be ideal for this purpose. XGBoost, on theother hand, is based on decision tree methods and is proven to be a highly efficient implementationof gradient-boosted decision trees. It is an ensemble learning technique that combines multipleweak decision tree models to create a strong model. Because it is based on decision tree methods, XGBoost is capable of learning complex, non-linear, and hierarchical relationships between variables. Thus, XGBoost is a suitable choice for this study.[7] Other machine learning methods areemployed as robustness checks.

Secondly, while coefficients or R-squares in linear regression models are easily interpretable, machine learning results can be quite challenging to understand. To address this issue, we utilize the SHAP method. Proposed by Lundberg and Lee (2017), SHAP employs game theory (Strumbelj and Kononenko, 2014) and local explanations (Ribeiro et al., 2016) to estimate the marginal contribution of each factor, measured using Shapley values. Shapley values are a measurement of an instance of a factor's average contribution across all possible coalitions and are used to determine each factor's contribution to the changes in a given outcome variable. Essentially, Shapley values help accurately allocate contributions to each factor, making machine learning results easier to understand. The visualization generated by Shapley values further simplifies thisprocess. For more information on SHAP, readers may refer to Huang (2022).

Thirdly, SHAP is an excellent tool to examine the interaction effects between factors, whichis one of the main objectives of this study. SHAP's visualization capabilities facilitate the understanding of these interactions.

[7] Readers may refer to Chen and Guestrin (2016) for more detailed information on XGBoost, particularly its mathematical aspects. Lundberg et al. (2020) demonstrate that gradient-boosted tree models - a more general version of XGBoost - can be both more accurate than other machine learning methods such as neural networks and more interpretable than standard linear models. Recent applications of XGBoost can be found in works by Huang (2022) and others.

8. Data and Results

8.1. Data

In this section, we introduce the data used in our study and present the results obtained through the machine learning approach. We use the RAND Health and Retirement Study (HRS)Longitudinal File 2018 as our data source. This file is a comprehensive and easy-to-use dataset containing public information (i.e., no restricted data) from the HRS Core and Exit interviews. It includes variables on a variety of topics, including demographics, health, health insurance, Social Security, pensions, retirement plans, family structure, employment history, as well as imputationsfor income, assets, and medical expenditures developed at RAND. All variables have consistent naming conventions and derivations across survey years, with any cross-wave differences documented.[8]

We use several variables from the dataset for our study, including stock and bond assets, age, self-reported health, earnings, and medical expenditures (out of pocket) for respondents. We compute the risky asset share as the ratio of stocks to the sum of stocks and bond assets for each respondent. Medical expenditure risk is calculated using Equation 1 through regression of recent out-of-pocket medical expenditures on the previous observation's expenditures for each individual.To perform our analysis, we only include respondents who have less than six missing values forstock, bond, and self-reported health variables and have less than three missing values for medical expenditures, where the number of positive medical expenditures is greater than two. Respondents who hold stock and bond assets in the next observation are excluded from the sample. Afterward,we calculate the variables' mean across all survey periods and retain only those with a mean age greater than 64 years (i.e., >= 65), resulting in a total sample of 1294 respondents.

Table 5 summarizes the statistics for the variables used in our machine learning analysis. The average risky share in our sample is 82%, indicating that most respondents hold more stock assets than bond assets. The average self-reported health is 2.3, which indicates that respondents' healthis generally good to very good.

Table 5. Summary Statistics.

	Risky Share (%)	HLT	OOP	Earning (log)
Min	0.23	1.00	0.17	-9.21
Mean	81.68	2.31	1.88	3.36
Median	90.43	2.00	0.97	8.19
Max	100.00	5.00	9.24	12.70
SD	21.72	0.79	2.00	8.32

Table 6. Regression results.

	Dependent variable:			
	ratio			
	(1)	(2)	(3)	(4)
OOP	0.07	0.04	0.05	-1.04
HLT		-2.42***	-2.31***	-3.30***
Earning			0.07	0.07
OOP: HLT				0.48
Constant	81.54***	87.19***	86.70***	88.96***
Observations	1,294	1,294	1,294	1,294
Adjusted R^2	−0.001	0.01	0.01	0.01

*Note: *p<0.1; **p<0.05; ***p<0.01.*

8.2. Results

In this section, we present the results of both regression analysis and machine learning techniques. Table 6 displays the regression results, indicating that medical expenditure risk is positively associated with risky share, but this relationship is not statistically significant when controlling for health and earnings. However, conditional on health, the association tends to change in size. Importantly, health has a statistically significant negative

8 Additional information on the RAND HRS Longitudinal File 2018 can be found at: https://hrsdata.isr.umich.edu/data-products/rand-hrs-longitudinal-file-2018.

association with risky share, even whencontrolling for the interaction term between OOP and HLT. Although the interaction term has an insignificant coefficient, we should be cautious in concluding that there are no interaction effects between medical expenditure risk and health, given that regression analysis relies on strongassumptions regarding the model specification. Thus, we also employ machine learning methods to examine the relationship between risky share, medical expenditure risk, and health, which does not impose any model specification.

Our machine learning results in Figure 1 indicate that health status is the most important variable in predicting risky share, and medical expenditure risk has less predictive power. This finding is consistent with our regression and structural model predictions, which also emphasize the significance of health. The SHAP values plotted in Figure 2 reveal that the higher the health insurance benefit, the lower the risky share (i.e., holding fewer stocks). Contrarily, respondents with lower medical expenditure risk tend to have a higher risky share. Furthermore, Figure 3 illustrates the interaction effects between health and medical expenditure risk. As health deteriorates, the SHAP values decrease, and respondents with good health tend to hold more safe assets when medical expenditure risk is high and invest in more stocks when medical risk is low.

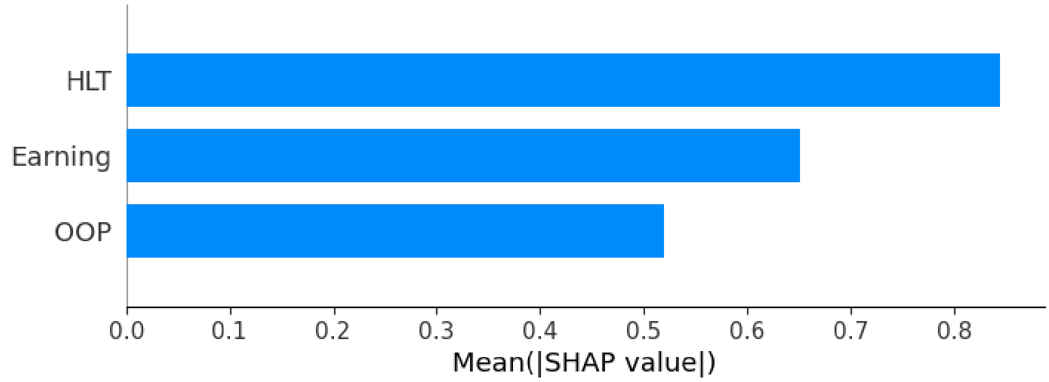


Figure 1. Average impacts on risky share.

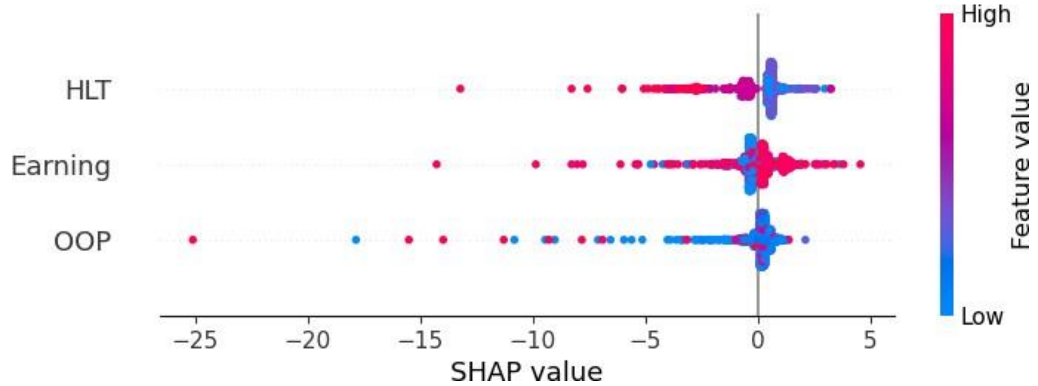


Figure 2. Impacts on risky share.

In summary, combining both regression analysis and machine learning techniques, we find substantial evidence that health is a crucial factor that affects risky share, while medical expenditurerisk has a secondary impact.

8.2.1. Robustness Checks

This section presents the robustness results of our study. First, we apply two common machine learning techniques, support vector machine, and neural networks, to replicate our analysis.Secondly, we balance our data and conduct the same research using XGBoost. We perform undersampling based on nearness to balance the data, as we divide households by HLT to good andpoor health groups in order to conduct some estimations, as shown in Table 1. Figures 4 and 5 display the results of our robustness checks, which are quite similar to the main findings. HLT is still the most important factor to predict risky shares and is negatively correlated with it.

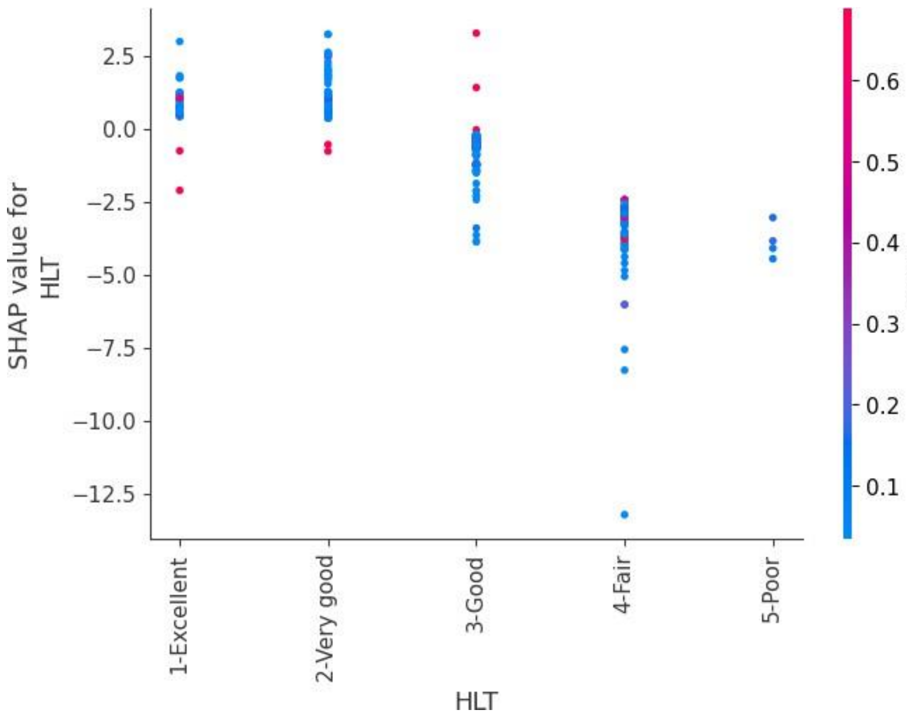
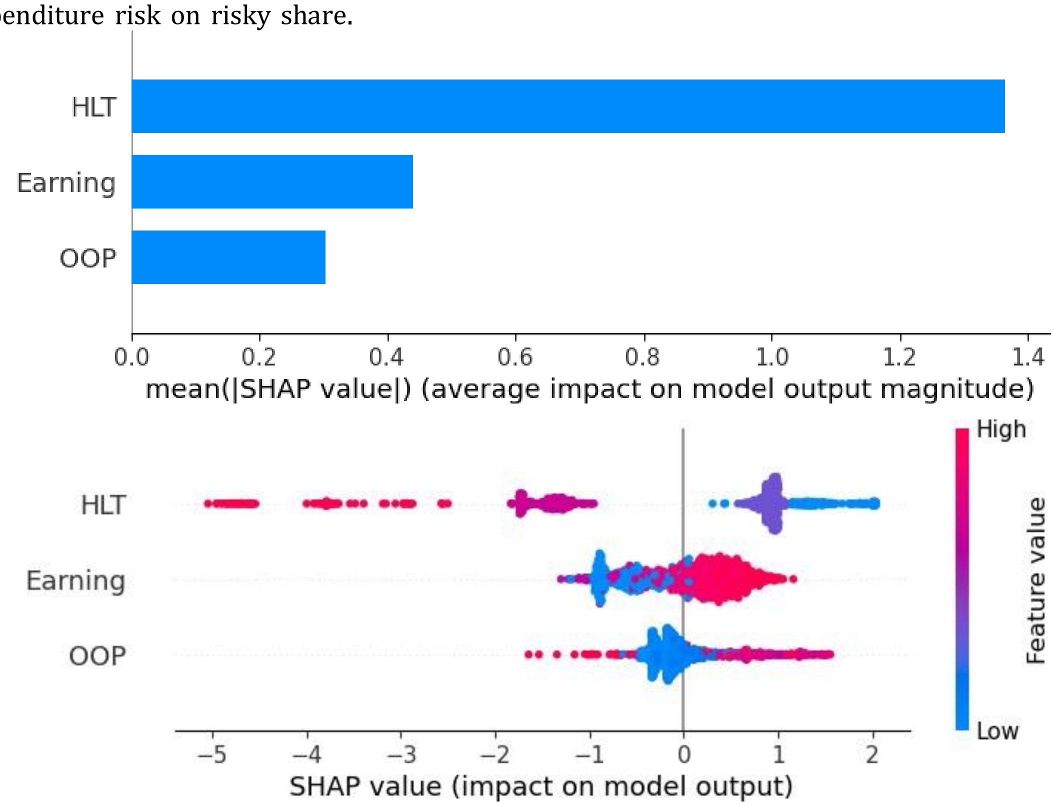


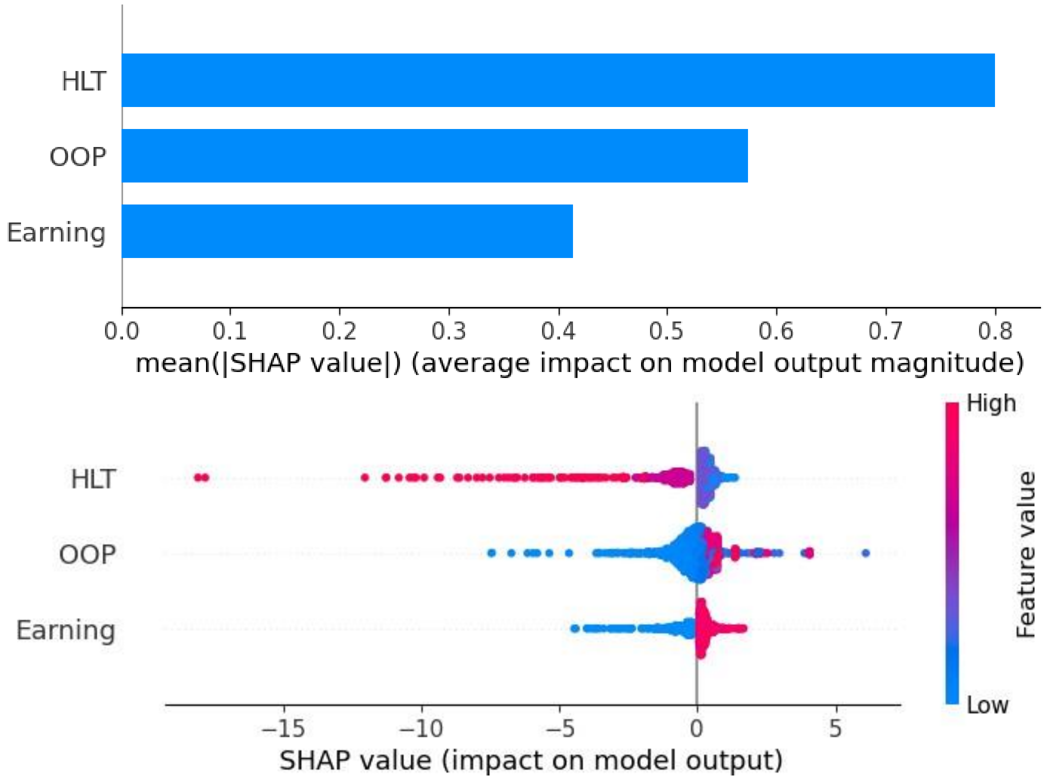
Figure 3. Interaction effects of health and medical expenditure risk on risky share.

Figure 6 presents the interaction effects between health and expenditure risk using supportvector machine and neural networks. As with XGBoost, we observe that as health deteriorates,risky share becomes smaller, consistent with our main results. Additionally, conditional on beingin good health, respondents with higher medical risk tend to have a lower risky share, which is consistent with the results obtained using XGBoost, as shown in Figure 3.

To conclude, our robustness checks using different machine learning techniques and a balanced data approach confirm the main results of the study, further strengthening our findings regardingthe impact of health and medical expenditure risk on risky share.



Support Vector Machine



Neural Networks

Figure 4. Average impacts on risky share using other ML methods.

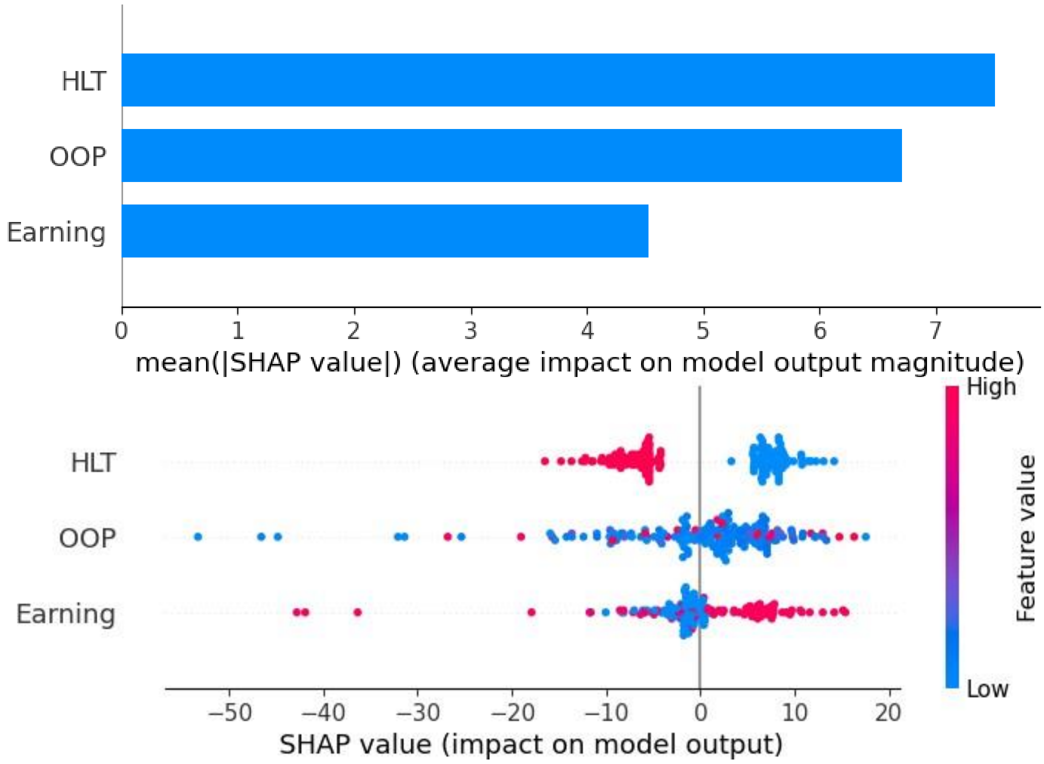
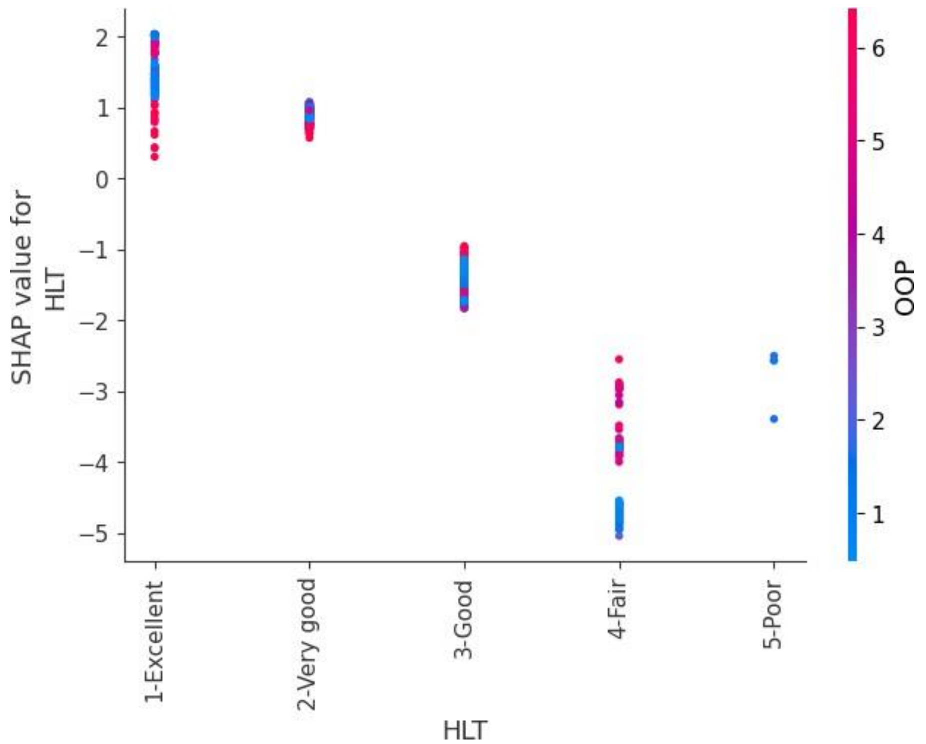
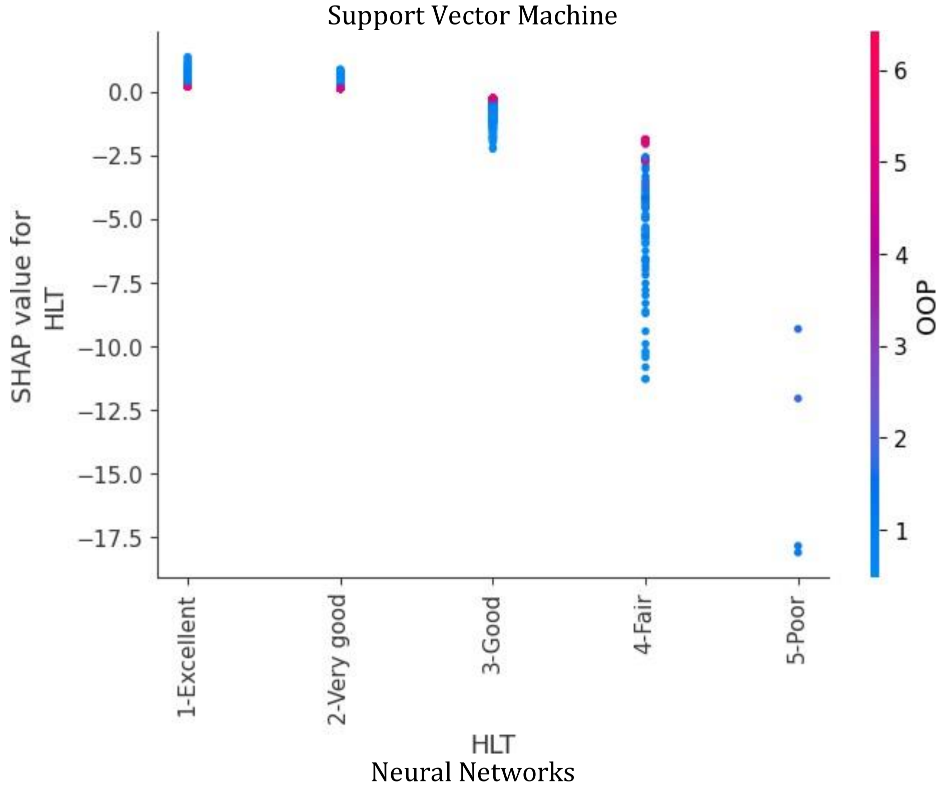


Figure 5. Average impacts on risky share using a balanced data.



Support Vector Machine



Neural Networks

Figure 6. Interaction effects of health and medical expenditure risk on risky share using other ML methods.

9. Conclusion

This paper provides a comprehensive analysis of the relationship between medical expenditure risk, health status, and portfolio choices. By integrating a life cycle model with machine learning techniques, we are able to examine both the theoretical predictions and empirical evidence regarding these relationships. The results of our study reveal several key findings. First, we find that relative to households with good health, the poor health group has a higher medical expenditurerisk and holds a portfolio with a less risky share. Additionally, when the medical expenditure risk is eliminated from the model, both the good health group and the poor health group increase their allocation to risky assets. This suggests that in the absence of medical expenditure risk, households are more inclined to invest in riskier assets, potentially seeking higher returns. However, it isworth noting that the effect is more pronounced in the poor health group, indicating that medicalexpenditure risk has a greater impact on their portfolio choices. To strengthen the robustness of our findings, we conduct various sensitivity analyses using alternative machine learning methodsand data balancing techniques.

Our results advance the understanding of the causal relationship between health, medical expenditure risk, and portfolio allocations. More specifically, the life profile analysis of medical expenditure risk in our paper fills the gap in this field and contributes to the literature. In addition, the cutting-edge methods of structural modeling and machine learning promote the existingmethodologies in household finance and health research fields. Overall, our findings contribute to the existing literature by shedding light on the mechanisms through which health and medical expenditure risk influence households' portfolio choices. By using innovative methodologies, we not only advance our understanding of this relationship but also provide valuable insights for policymakers and practitioners in the areas of household finance and healthcare.

However, our study has some limitations. Firstly, we focus on only health variables, while otherfactors like education, and demographic characteristics could also impact households' portfolio choices. Secondly, our study is limited to the US health care system, and the results may not be generalizable to other countries.

Our study's policy implications are clear. Policymakers and practitioners should be aware of the significant impact of medical expenditure risk on households' portfolio choices, especially forthose with poor health. They should also consider ways to mitigate medical expenditure risk and increase households' access to affordable health care, especially for low-income households. Finally, our study's innovative analytical techniques could be useful for developing more targeted policiesand interventions, and addressing the important issue of how health and medical expenditure riskinfluence households' financial decisions.

Funding Statement

This research is supported by "Fort Hays State University Faculty Summer Research Grant (2022)".

Acknowledgments

The authors are grateful to anonymous referees and the journal's editor for their helpful comments and efforts in improving this paper.

Conflict of interest

All the authors claim that the manuscript is completely original. The authors also declare no conflict of interest.

References

Ayyagari, P. and D. He. (2017). The Role of Medical Expenditure Risk in Portfolio Allocation Decisions. *Health Economics*, 26 (11), 1447-1458. https://doi.org/10.1002/hec.3437

Ban, G. Y., El Karoui, N., and Lim, A. E. (2018). Machine learning and portfolio optimization. *Management Science*, 64(3), 1136-1154. https://doi.org/10.1287/mnsc.2016.2644

Brunnermeier, M. K. and S. Nagel. (2008). Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-Evidence on Individuals' Asset Allocation. *The American Economic Review*, 98 (3), 713-736. https://doi.org/10.1257/aer.98.3.713

Campanale, C., C. Fugazza and F. Gomes. (2015). Life-cycle portfolio choice with liquid and illiquid financial assets. *Journal of Monetary Economics*, 71, 67-83. https://doi.org/10.1016/j.jmoneco.2014.11.008

Capatina, E. (2015). Life cycle effects of health risk. *Journal of Monetary Economics*, 74, 67-88. https://doi.org/10.1016/j.jmoneco.2015.06.002

Capatina, E., M. Keane, S. Maruyama. (2018). Health Shocks and the Evolution of Consumption and Income over the Life-Cycle. Working paper.

Capatina,E., M. Keane, S. Maruyama. (2018). Socio-economic Disparities in US Healthcare Spending: The Role of Public vs Private Insurance. UNSW Business School Paper No.2018-03.

Castello-Climent, A and R. Domenech. (2008). Human Capital Inequality, Life Expectancy and Economic Growth. *The Economic Journal*, 118 (528), 653-677. https://doi.org/10.1111/j.1468-0297.2008.02136.x

Chang, Y., J. H. Hong and M. Karabarbounis. (2018). Labor Market Uncertainty and Portfolio Choice Puzzles. *American Economic Journal: Macroeconomics*, 10 (2), 222-62. https://doi.org/10.1257/mac.20160207

Chen, T., and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, [In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794)]. https://doi.org/10.1145/2939672.2939785

Chiappori, P. and M. Paiella. (2011). Relative Risk Aversion is Constant: Evidence from Panel Data. *Journal of the European Economic Association*, 9 (6), 1021-1052. https://doi.org/10.1111/j.1542-4774.2011.01046.x

Cocco, J.F., F. Gomes and P.J. Maenhout. (2005). Consumption and Portfolio Choice over the Life Cycle. *The Review of Financial Studies*, 18 (2), 491-533. https://doi.org/10.1093/rfs/hhi017

Cocco,J.F. (2005). Portfolio Choice in the Presence of Housing. *The Review of Financial Studies*, 18 (2), 535-567. https://doi.org/10.1093/rfs/hhi006

Deaton, A. (2003). Health, Inequality, and Economic Development. *Journal of Economic Literature,* 41(1), 113-158. https://doi.org/10.1257/002205103321544710

De Nardi, M., E. French and J. Jones. (2010). Why do the elderly save? The role of medical expenses. *The Journal of Political Economy*, 118 (1), 39-75. https://doi.org/10.1086/651674

De Nardi, M., S. Pashchenko and P. Porapakkarm. (2018). The Lifetime Costs of BadHealth. Working paper.

Du, Y. (2021). Optimal Consumption and Portfolio Choice: The Impact of Health andHealth Risk. *The Economics and Finance Letters*, 8(2), 266–276. https://doi.org/10.18488/journal.29.2021.82.266.276

Edwards, R.D. (2008). Health Risk and Portfolio Choice. *Journal of Business & EconomicStatistics*, 26 (4), 472-485. https://doi.org/10.1198/073500107000000287

Fagereng, A., C. Gottlieb, L. Guiso. (2017). Asset Market Participation and Portfolio Choiceover the Life-Cycle. *The Journal of Finance*, 72 (2), 705-750. https://doi.org/10.1111/jofi.12484

Fagereng, A., L. Guiso and L. Pistaferri. (2018). Portfolio Choices, Firm Shocks, and Uninsurable Wage Risk. *The Review of Economic Studies*, 85 (1), 437-474. https://doi.org/10.1093/restud/rdx023

Fan, X., A. Seshadri and C. Taber. (2015). Estimation of a Life-Cycle Model with Human Capital, Labor Supply and Retirement. Working paper.

Fuchs, Victor R. (1966). The Contribution of Health Services to the American Economy. *The Milbank Memorial Fund Quarterly*, 44(4), 65-103. https://doi.org/10.2307/3349060

Goldman, D. and N. Maestas. (2013). Medical expenditure risk and household portfolio choice. *Journal of Applied Econometrics*, 28(4), 527-550. https://doi.org/10.1002/jae.2278

Gomes, F and A. Michaelides (2005). Optimal life cycle asset allocation: understanding the empirical evidence. *The Journal of Finance*, 60 (2), 869-904. https://doi.org/10.1111/j.1540-6261.2005.00749.x

Grossman, M., (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy*, 80 (2), 223-255. https://doi.org/10.7312/gros17812-004

Guiso, L. and M. Paiella. (2008). Risk Aversion, Wealth, and Background Risk. *Journalof the European Economic Association*, 6 (6), 1109-1150. https://doi.org/10.1162/JEEA.2008.6.6.1109

Halliday, T.J., H. He, L. Ning, and H. Zhang. (2019). Health Investment over the Life-Cycle. *Macroeconomic Dynamics*, 23, 178-215. https://doi.org/10.1017/S1365100516001152

Hall, R.E. and C.I. Jones. (2007). The Value of Life and the Rise in Health Spending. *The Quarterly Journal of Economics*, 122 (1), 39-72. https://doi.org/10.1162/qjec.122.1.39

He,H., K.X.D. Huang and L. Ning. (2019). Why Do Americans Spend So Much More on Health Care than Europeans?. Vanderbilt University Department of Economics Working Papers.

Heaton, J. and D. Lucas. (1996). Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing. *Journal of Political Economy*, 104(3), 443-487. https://doi.org/10.1086/262030

Heaton, J. and D. Lucas. (2000). Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk. *The Journal of Finance*, 55 (3), 1163-1198. https://doi.org/10.1111/0022-1082.00244

Hosseini, R., K. A. Kopecky and K. Zhao. (2019). The Evolution of Health over the Life Cycle. Working paper.

Hosseini, R., K.A. Kopecky and K. Zhao. (2020). How Important is Health Inequality for Lifetime Earnings Inequality?. Working paper.

Huang, W. (2022). Digesting Three-factor Model. *The Singapore Economic Review*, 1-30. https://doi.org/10.1142/S0217590822480022

Huggett, M. and G. Kaplan. (2016). The Money Value of a Man. Working paper.

Inkmann, J., Lopes, P., and Michaelides, A. (2011). How deep is the annuity market participation puzzle?. The Review of Financial Studies, 24(1), 279-319.

Jayachandran, S. and A. Lleras-Muney. (2009). Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines. *The Quarterly Journal of Economics*, 124(1), 349-397. https://doi.org/10.2307/40506231

Kaushik, S., Choudhury, A., Natarajan, S., Pickett, L. A., and Dutt, V. (2020). Medicine expenditure prediction via a variance-based generative adversarial network. *IEEE Access*, 8, 110947-110958. https://doi.org/10.1109/ACCESS.2020.3002346

Khwaja, A. (2010). Estimating willingness to pay for medicare using a dynamic life-cycle model of demand for health insurance. *Journal of Econometrics*, 156(1), 130-147. https://doi.org/10.1016/j.jeconom.2009.09.011

Kimball, Miles S. (1990). Precautionary Saving in the Small and in the Large. *Econometrica*, 58(1), 53-73. https://doi.org/10.2307/2938334

Koijen, R.S.J., S. Van Nieuwerburgh, and M. Yogo. (2016). Health and mortality delta: assessing the welfare cost of household insurance choice. *The Journal of Finance*, 71 (2), 957-1010. https://doi.org/10.1111/jofi.12273

Kopecky, K. A. and T, Koreshkova. (2014). The impact of Medical and Nursing Home Expenses on Savings. *American economic journal: Macroeconomics*, 6 (3), 29-72. https://doi.org/10.1257/mac.6.3.29

Lockwood, L. M. (2018). Incidental Bequests and the Choice to Self-Insure Late-Life Risks. *American Economic Review*, 108(9), 2513-50. https://doi.org/10.1257/aer.20141651

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... and Lee, S. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56-67. https://doi.org/10.1038/s42256-019-0138-9

Michaud, A and D. Wiczer. (2017). The Disability Option: Labor Market Dynamics with Macroeconomic and Health Risks. Department of Economics Working Papers 18-12, Stony Brook University, Department of Economics.

Mushkin, Selma J. (1962). Health as an Investment. *Journal of Political Economy*, 70, 129. https://doi.org/10.1086/258730

Nagarjuna, M., Pasula, P., Kavyakeerthi, T., and Karthik, I. (2022). Medicine Expenditure Prediction using Machine Learning, *In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC)* pp. 948-952, IEEE. https://doi.org/10.1109/ICCMC53470.2022.9753780

Nakajima, M. and I.A. Telyukova. (2018). Medical Expenses and Saving in Retirement:The Case of U.S. and Sweden. Institute Working Paper 8.

Nakajima, M. and D. Tüzemen. (2017). Health Care Reform or Labor Market Reform? A Quantitative Analysis of the Affordable Care Act. Working paper.

Oster, E., I. Shoulson and E. R. Dorsey. (2013). Limited Life Expectancy, Human Capitaland Health Investments. *The American Economic Review*, 103(5), 1977-2002. https://doi.org/10.1257/aer.103.5.1977

Ozkan, S. (2017). Preventive vs Curative Medicine: A Macroeconomic Analysis of Health Care. Working paper.

Pang, G.and M. Warshawsky. (2010). Optimizing the equity-bond-annuity portfolio in retirement: The impact of uncertain health expenses. *Insurance, Mathematics* & E*conomics*, 46 (1), 198-209. https://doi.org/10.1016/j.insmatheco.2009.08.009

Papageorge, N. W. (2016). Why medical innovation is valuable: Health, human capital, and the labor market. *Quantitative Economics*, 7(3), 671-725. https://doi.org/10.3982/QE459

Paravisini, D., V. Rappoport, and E. Ravina. (2017). Risk Aversion and Wealth: Evidence from Person-to-Person Lending Portfolios. *Management Science*, 63 (2), 279-297. https://doi.org/10.1287/mnsc.2015.2317

Prados, M.J. (2012) Health and earnings inequality over the life cycle and the redistributivepotential of health policies. Working paper.

Qin, F. Y., Lv, Z. Q., Wang, D. N., Hu, B., and Wu, C. (2020). Health status prediction forthe elderly based on machine learning. *Archives of Gerontology and Geriatrics*, 90, 104121. https://doi.org/10.1016/j.archger.2020.104121

Rosen, H.S. and S. Wu. (2004). Portfolio choice and health status. *Journal of Financial Economics*, 72 (3), 457-484. https://doi.org/10.1016/S0304-405X(03)00178-8

Shapley, L. S. (1997). A value for n-person games. *Classics in Game Theory*, 69.

Tarekegn, A., Ricceri, F., Costa, G., Ferracin, E., and Giacobini, M. (2020). Predictive modeling for frailty conditions in elderly people: machine learning approaches. JMIR Med ical Informatics, 8(6), e16678. https://doi.org/10.2196/16678

Wachter, J.A. and M. Yogo. (2010). Why do household portfolio shares rise in wealth?. *The Review of Financial Studies*, 23 (11), 3929-3965. https://doi.org/10.1093/rfs/hhq092

White, M.N. (2018). Optimal Health Investment: An Ounce of Prevention at Half Price?. Working paper.

Yao, R. and H. H. Zhang. (2005). Optimal consumption and portfolio choices with riskyhousing and borrowing constraints. *The Review of Financial Studies*, 18 (1),197-239. https://doi.org/10.1093/rfs/hhh007

Yogo, M. (2016). Portfolio choice in retirement: Health risk and the demand for annuities,housing, and risky assets. *Journal of Monetary Economics*, 80, 17-34. https://doi.org/10.1016/j.jmoneco.2016.04.008