Information Engineering Express
International Institute of Applied Informatics
2023, Vol. 9, No. 1, IEE693

Further Analysis of Anonymous Data to Assess Dissaving Risk against Life Expectancy for Elderly People

Yuya Yokoyama *, Yasunari Yoshitomi *

Abstract

No database exists suitable for disclosing the effect on economic activities caused by dementia. Thus, our aim is to obtain the principles to resolve its influence by analyzing related databases. In order to detect the capability deterioration of economic activity for elderly people at the age of sixty-five years or older, we have employed an anonymous data set obtained from the National Survey of Family Income and Expenditure carried out by Ministry of Internal Affairs and Communications. So far, a method has been developed to detect dissaving risk of elderly people using clustering. Nevertheless, the initial approach showed limitations for further improvement. Therefore, we have proposed another method to detect dissaving risk by dividing analysis data into test/training data. The test data are then processed through primary/secondary determination in accordance with the amount of savings. In secondary determination, three methods are performed in terms of income and savings. In the previous analysis, as an initial step, merely the data of single females were analyzed. In this paper, the data of single males and two-person households are analyzed. As a result of the analysis, the correct judgment rate exceeds 80% for single males, while it attains as high as approximately 70% for two-person households.

Keywords: Anonymous data, Dissaving risk, Life expectancy, Savings

1 Introduction

To the best of our knowledge, there has been no existing database suitable for unraveling the influence on economic activities caused by dementia. Therefore, in researching the influence on economic activities resulting from dementia, we have set out by analyzing the existing database related to economic activities of elderly people. The decreasing nerve cells and the decline of cognitive function along with aging are universally known as causing the deterioration of memory, thinking, and motivation [1]. Cerebral disorder is also known by the symptoms of dementia, e.g. memory impairment, judgment disorder, execution function disorder, dropping off of thinking ability, etc. It has also been known that aging can increase the possibility of causing the onset of the disease. Symptoms of depression can also deprive patients of their vitality and even interest in what they used to be interested in before.

* Kyoto Prefectural University, Kyoto, Japan

There are an enormous number of cases where temporal, economic, and geographical disparities form obstacles to early diagnosis of dementia by doctors. Thanks to several existing technologies, dementia can be determined by analyzing the pattern of their behavior in the daily lives of the elderly single persons. For instance, a system has been developed that estimates dementia through the usage of input information in order to inform and input their safety [2]. This system makes it possible to observe the safety under the protection of the elderly's privacy as well as foresee dementia without regular medical check-ups. Another example is a household electric appliance that determines if the operator is suffering from dementia by retrieving his/her operation history, sleeping hours, commodity purchase history, etc [3]. If the possibility of dementia is observed and determined, the information is sent to an external device possessed by the family of the operator. Nevertheless, it would require permission to monitor the behavior of elderly people to determine dementia. Therefore, such devices must be purchased and installed with a view to monitoring the intended party.

Jeopardized economic activity can be divided into the following two types: (1) "Dissaving risk," including the risk of becoming victimized by fraudulent activity, which results from the deterioration of judgment, etc., (2) "Under-consumption risk," which is caused by the lowered motivation to spend money. Therefore, our objective is to develop a system that detects such jeopardized economic activity early and issues a warning. In this paper, the first, "dissaving risk" is focused on and analyzed.

So far, in our initial method, we used an anonymous data on the basis of the National Survey of Family Income and Expenditure (NSFIE) [4] carried out by the Ministry of Internal Affairs and Communications (MIC) in 2004. As a basis of economy activity, for the data of single-person households (Singles) at the age of sixty-five or older, clustering was performed with five clusters based on income and saving [5]. As a result, five clusters were obtained and characteristically classified. Then we developed a method to determine if savings were sufficient against life expectancy in terms of income and savings [5]. Consequently, a correction detection ratio of the data was determined as having a dissaving risk that ranged from 52.6 to 94.1%.

This previous method, however, could not be developed further with the addition of the dataset for the years 1994 and 1999. Therefore, we have shifted to propose another method to detect dissaving risk of people at the age of sixty-five or older. Here, the analysis data are divided into test data and training data [6]. Two thresholds regarding savings are then determined. One is used for determining whether the test data will be further processed in either one of two types of determination. The other is utilized to set a training data. If one test data is determined through secondary determination, three methods are then performed in terms of income and savings. As an initial step of the method, the data of females were analyzed [6]. The analysis result has shown that the correct judgment rate (CJR) exceeded 80% for single females.

So far merely the data of single females have been analyzed for the proposed methods [6]. Therefore, in this paper we proceed to analyze the rest of the analysis data, that is, single males and two-person households, in a similar fashion as the case of single females. As a result of the further analysis, CJR outnumbers 80% for single males as well, whereas it only reached as high as approximately 70% for two-person households.

The remainder of the paper is organized as follows. Related works are explained in Section 2. Preparation of the analysis is summarized in Section 3. In Section 4, the methodology to evaluate our proposed method to determine dissaving risk is stated. The evaluation of the proposed method is provided in Section 5. Finally, Section 6 concludes the paper.

2 Related Works

According to Fujita *et al.*, the reduction of economic activities for rapidly increasing dementia patients in accordance with aging could have a bad influence on the actual economy [7]. Taking an example of an elderly person suffering from dementia, he/she might repetitively act out the following experiences: he/she fails to withdraw money from his/her bank account because of not memorizing the password, forgets to pay bills, repeatedly purchases the commodities he/she already had before. These experiences could result in considerable reduction of economic activities. In the 2012 fiscal year, among elderly people at the age of sixty-five or older, there are roughly 4.62 million dementia patients, whereas there are approximately 4 million with mild cognitive impairment (MCI) [7]. The total number of these patients is equivalent to approximately 7% of the Japanese population. Since category of dementia patients is mostly occupied by elderly people, their number is expected to keep increasing on a global scale in the upcoming several decades [7]. Therefore, the shrinking of economic activities will have an extreme impact on the economy worldwide. From these future perspectives, we focus on the possible deterioration of economic activities of elderly people.

Agarwal *et al*. made an analysis of cognitive ability by employing the actual examples of almost best behaviors and ample economic decisions of consumers [8]. The analysis results conveyed the conclusion that consumers with overall higher scores on analysis items and mathematics are unlikely to end up with economic collapse. Bisdee *et al.* investigated money management issues of the elderly households along with aging and cognitive ability decline [9]. For the purpose of detecting the evidence of intelligence decline in accordance with aging, Finke *et al.* revealed that if the knowledge of basic concepts necessary for effective and economical choice were to deteriorate after the age of sixty years, that could influence the ability necessary for money management [10]. Hsu *et al.* analyzed the impact of information against cognitive ability deterioration of decision maker of finance by employing time-series data related to elderly married couples [11]. These research studies [8-11] were conducted based on anecdotal reports and multiple manifold methods, e.g. regression analysis. However, there has been no research focusing on the relationships between cognitive ability and purchase motivation.

3 Preparation for Analysis

3.1 Anonymous Data (AD)

The analysis data used for our work is an anonymous data (AD) obtained from the National Statistics Center [12]. The AD is based on the NSFIE carried out by the MIC in 1994, 1999 and 2004. These datasets contain 1,919 (1,752 and 1,780, respectively) detailed records of family income and expenditure survey items for 1994 (1999 and 2004), e.g. the amount of income and savings, expenditure amount of food such as grain, fish, etc. Among the AD, this paper uses the data of Singles and Twos at the age of sixty-five or over; 717 single males, 3,475 single females, and 11,528 Twos.

Because the latest year of AD issued by the MIC is 2004, we must consider whether the economic situation of 1994, 1999, and 2004 on the basis of AD would not deviate from those of

2004, 2009 and 2014 based on the NSFIE [13]. In order to compare those economic situations, the amounts of data, average income, average savings, and Consumer Price Index (CPI) of Singles at the age of 65 or older in 1994, 1999, 2004, 2009 and 2014 for males are shown in Table 1. The results of NSFIE pressed by MIC are shown in the columns entitled 2004, 2009, and 2014. On the other hand, those of AD that include a portion of NSFIE are provided in the column entitled 1994, 1999, and 2004. For this reason, the data of NSFIE for 2004 do not correspond to those of AD for 2004. Both income and savings are provided as units of ten thousand yen per year. Each CPI in 1994, 1999, 2004, 2009, and 2014 is expressed with the criteria that CPI in 2015 is set to 100 [14]. From the viewpoints of income, savings and CPI shown in Table 1, it could be suggested that the economic situation in 1994, 1999, and 2004, which is used in our work, does not considerably differ from that in 2014, which is relatively closer to the current 2021 [15].

Table 1: Comparison of Economic Situation [13]

(a) Male

Item	1994	1999	2004		2009	2014
	AD	AD	NSFIE	AD	NSFIE	NSFIE
Data	317	229	368	317	359	593
Income*	272.1	273.0	285.8	277.6	275.4	260.7
Savings*	1328.6	1294.5	1816.1	1313.7	1265.2	1501.8
CPI	97.7	99.8	97.2		97.2	99.2

* 10 thousand yen

(b) Female

Item	1994	1999	2004		2009	2014
	AD	AD	NSFIE	AD	NSFIE	NSFIE
Data	957	1133	1577	1385	1606	2123
Income*	203.4	215.8	229.0	219.0	243.5	218.5
Savings*	962.9	1059.3	1422.4	1131.3	1594.4	1466.2
CPI	97.7	99.8	97.2		97.2	99.2

* 10 thousand yen

3.2 Revised Disbursement (Re-dis)

Before analysis, the contents of disbursement must be carefully regarded. Disbursement is composed of the following three components: (1) Expenditure, (2) Disbursements other than expenditure, and (3) Carry-over to next. Among these elements, we have extracted the items that seem doubtful for being considered as disbursement [5]. From this viewpoint, five items were excluded from disbursement [5].

As explained in Section 3.1, income is expressed in ten thousand yen units per year in AD. On the other hand, each household item related to disbursement is provided by yen per month. Therefore, the value of disbursement, exempting the five items above, is multiplied by 12 to calculate yearly costs. This value is defined as "Revised disbursement" (Re-dis). The amount of revised disbursement is denoted as D_{rev} Using Formula (1), D_{rev} is calculated [5].

$$D_{rev} = \frac{12\{D_1 + (D_2 - v_1 - v_2 - v_3 - v_4)\}}{10000} \tag{1}$$

,where D_1 is the amount of expenditure, D_2 is the total amount of disbursements other than expenditure, v_1 is the saving, v_2 is the amount of insurance premium payments, v_3 is that of security purchase, and v_4 is that of property purchase.

The terms v_1, v_2, v_3, v_4 are the feature values exempted from disbursement. The unit of D_{rev} is ten thousand yen, while those of $D_1, D_2, v_1, v_2, v_3, v_4$ are yen. In order to unify the unit yen, the right side of Formula (1) is divided by 10,000.

3.3 Selection of the Analysis Data among the AD

Among the AD, in order to consider a method to determine if savings will be ample against life expectancy, the following two types of data are exempted from the subsequent analysis. Here, Re-dis is set as D_{rev} ,while income is set as I.

(1) The data whose Re-dis is not more than income ("$D_{rev} \leq I$") are exempted, since their savings are unlikely to be decreased.

(2) The data whose Re-dis is more than income ("$D_{rev} > I$") and savings are 0 yen are also excluded, because no warnings would be urged for their savings.

Therefore, the analysis data used for the following analysis are AD meeting both "$D_{rev} > I$" and some savings. The amounts of the analysis data for Single females are shown in Table 2.

Table 2: Amounts of Analysis Data

Year	Single						Two-person Household		
	Male			Female					
	Rent	Own	Total	Rent	Own	Total	Rent	Own	Total
1994	11	30	41	104	234	338	78	445	523
1999	24	31	55	105	275	380	96	613	709
2004	26	49	75	122	363	485	132	1003	1135
Total	61	110	171	331	872	1203	306	2061	2367

4 Methodology to Determine Dissaving Risk

4.1 Requirement and Outline of Proposed Method

In our initial method, we merely employed AD dated to 2004. As a basis of economic activity, for the data of Singles at the age of sixty-five or older, clustering was performed with five clusters based on income and saving [15]. The analysis results contributed to obtaining five clusters that were characteristically classified. This first method was then developed to determine if savings were ample against life expectancy in terms of income and savings [15]. Consequently, the correction detection ratio of the data determined as having a dissaving risk ranged from 52.6 to 94.1%.

However, the initial method showed the limitations with the supplement of the AD dated to 1994 and 1999. Therefore, we have altered our course by proposing another method to determine if savings are sufficient against life expectancy in terms of savings. As an initial step of this proposed method, only the data of single females were analyzed [5]. Here, the term in which savings are exhausted, which is defined as "saving exhaustion term" (SET), is utilized to determine "dissaving risk". If the figure is less than the threshold against life expectancy, the data are determined as having "dissaving risk". For the analysis data chosen in Section 3.3, rent and house are separately considered and analyzed. For the test data randomly extracted from the analysis data, the conditions as to savings, income and Re-dis are regarded. The data meeting these conditions are utilized as training data. Additionally, the thresholds of savings per age group are set.

The outline of evaluating the proposed method through randomly extracting test data from the data analysis is summarized as follows. Firstly, as primary determination, the data whose savings are not less than threshold is determined as having "no dissaving risk." Next, as secondary determination, the rest of data are evaluated and compared through three types of analytical methods, in terms of income and savings. Here, the three data are extracted from the partial sets out of the training data according to the order of distance among the vector space composed of income and savings. Among the three SETs in which savings will be exhausted, the shortest one is set as "estimated term". The estimated term of a test data is then divided by life expectancy corresponding to its age group and survey year. If the figure obtained is less than 0.1 (0.3, respectively) for Single (Two), the data is determined as "dissaving risk".

4.2 Condition to Determine Dissaving Risk

4.2.1 Representative Age and Life Expectancy

Assuming that the difference between Re-dis and income is constant, the SETs are obtained through dividing the difference by savings. The term is described as T. T is obtained through Formula (2) [5].

$$T = \frac{S}{D_{rev} - I} \tag{2}$$

, where S is the amount of savings, I is that of income, and D_{rev} is that of Re-dis.

Dissaving risk is evaluated using life expectancy (L_{exp}) in 1994, 1999, and 2004 published by the Ministry of Health, Labor and Welfare (MHLW) [13]. In the AD, age is provided as a form of age group. Actual age for each data is completely unknown. Among the age groups, the youngest age is defined as "representative age" and denoted as A_{rep}. Taking an example of ages 65 to 69, A_{rep} is 65. The L_{exp} for 1994, 1999 and 2004 is then used in the form of a figure rounded off to the nearest integer. L_{exp} for each age group is shown in Table 3.

For A_{rep} and L_{exp} of Twos, these figures are averaged. Taking an example of a household surveyed on 1999 and composed of a male at the age of 70 to 74 whose L_{exp} is 13 years old and a female at the age of 65 to 69 whose L_{exp} is 22 years old, each A_{rep} is 70 and 65, respectively. Their average A_{rep} is 67.5, whose A_{rep} is regarded as 65 since the number is between 65 and 70. Their average L_{exp} is 17.5 years. L_{exp} for Twos are listed in Table 4.

Table 3: Life Expectancy with the Assumption of Youngest in Age Group

Age Group	A_{rep}	Male			Female		
		1994	1999	2004	1994	1999	2004
65 - 69	65	17	17	18	21	22	23
70 - 74	70	13	13	15	17	18	19
75 - 79	75	10	10	11	13	14	15
80 - 84	80	7	8	8	9	10	11
85 -	85	5	5	6	7	7	8

Table 4: Life Expectancy of Two-person Household

		Female				
	Age (A_{rep})	65 - 69 (65)	70 - 74 (70)	75 - 79 (75)	80 - 84 (80)	85 - (85)
Male	65 - 69 (65)	65	65	70	70	75
	70 - 74 (70)	65	70	70	75	75
	75 - 79 (75)	70	70	75	75	80
	80 - 84 (80)	70	75	75	80	80
	85 - (85)	75	75	80	80	85

4.2.2 Condition of Dissaving Risk Determination

Given that P is the ratio of T to L_{exp} corresponding to its age category for each data, P is calculated through Formula (3).

$$P = \frac{T}{L_{exp}} \tag{3}$$

Provided that P_{th} is the threshold to determine dissaving risk of the data, if P is less than P_{th}, the data is determined as having dissaving risk. In this paper, P_{th} is set to 0.1 (0.3, respectively) for Singles (Twos) through trial and error.

4.3 Dividing Analysis Data into Training/Test Data

4.3.1 Training/Test Data

The analysis data selected in Section 3.3 are divided into the following two groups: “rent” and “own”. In this paper, rent and own are separately considered and analyzed. Here, the analysis data are divided into six groups of attributes consisting of house types (“Rent” and “Own”) and year (“1994,” “1999” and “2004”). In order to investigate the validity of our proposed method, these data, classified into the respective six groups, are divided into training data and test data.

4.3.2 Training Data and Partial Sets

In order to select training data for the analysis data set in Section 3.3, the following three conditions are determined.

- Con_1: The amount of savings is less than or equal to the threshold S_{th} for A_{rep} corresponding to the age group of the training data.
- Con_2: Re-dis is 1.2 (1.15, respectively) times or more than income for Singles (Twos)
- Con_3: Saving is 0.1 times or less than income.

The figures 1.2, 1.15, and 0.1 are set through trial and error. The set that satisfies Con_1 (Con_2 and Con_3, respectively) is denoted as A (B and C). Con_1 is imposed because a little amount of savings could result in higher jeopardy of dissaving risk. If the data can be deduced as mandatory money owing to specific reasons, e.g. renovations, purchasing new furniture, etc., those data are excluded from determining S_{th} for Con_1. How to set two thresholds S_{th} and S_{th0} will be explained in Section 4.4. Con_2 is set since excessive over-expenditure could lead directly to dissaving risk. Con_3 is also contrived since an over-expenditure tendency could be speculated with its higher possibility of dissaving risk if savings is much less than the total amount of income.

Set A is used as the whole set of training data. For sets B and C, the partial sets D_1, D_2 of the training data are then defined as Formulas (4) and (5) [5].

$$D_1 = A \cap (B \cup C) \tag{4}$$

$$D_2 = A \cap (\bar{B} \cap \bar{C}) \tag{5}$$

4.4 Methodology

4.4.1 Regression Line to Determine Training Data

How to set two thresholds of savings (S_{th0} and S_{th}) is explained by using an example of single females living in their own house [6]. The threshold S_{th0} is used to determine if the test data will be processed through either primary determination or secondary determination which is subsequently explained in Section 4.4.2. On the other hand, threshold S_{th} is employed to set a training data in a fashion explained in this section. For each group age, representative age A_{rep} and maximum saving S_{MAX} are shown in Table 5. Each value of S_{MAX} is obtained as the maximum savings for the respective A_{rep} for the data determined as having dissaving risk among the analysis data. Here, S_{MAX} for each A_{rep} is plotted as a brown square in Figure 1. The horizontal axis indicates A_{rep}, while the vertical axis shows S_{MAX}. Regression line R_l is shown as dashed line in Figure 1. The line R_l is shown as Formula (6) [6].

$$R_l: S_{MAX} = -41.9A_{rep} + 3665.3 \tag{6}$$

The vertical value obtained through R_l at each A_{rep} is defined as predictive value S_{th0}.

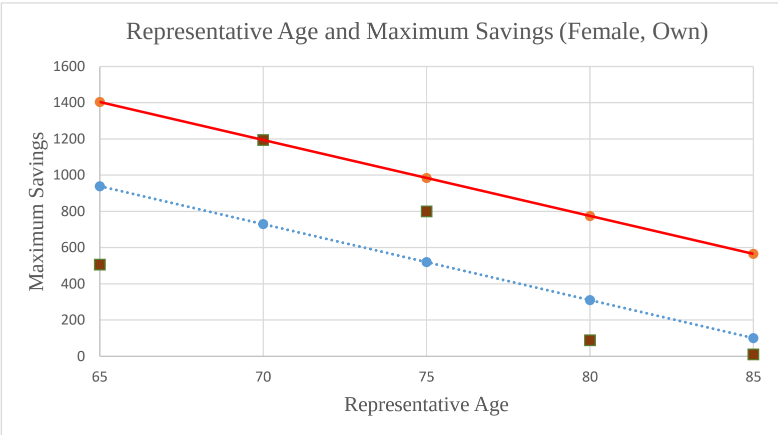


Figure 1: Example of Representative Age and Maximum Savings [5]

Provided that ΔS_{th} is defined as the predicted error obtained through subtracting S_{MAX} from S_{th0}, ΔS_{th} is expressed as the Formula (7).

$$\Delta S_{th} = S_{MAX} - S_{th0} \tag{7}$$

The whole set of training data *A* is defined as "the set whose savings are not more than those of the line R_l' defined as the vertically upward shift of R_l by the maximum of ΔS_{th} at the A_{rep} of its range focused on" [5]. R_l' is shown as a solid line in Figure 1 and as Formula (8) [6].

$$R_l': S_{MAX} = -41.9A_{rep} + 4129.8 \tag{8}$$

The maximum of ΔS_{th} is 464.5, which is shaded in Table 5. The vertical value obtained through line R_l' at each A_{rep} is defined as threshold S_{th}. On top of S_{MAX}, three values (S_{th0}, ΔS_{th} and S_{th}) are also shown in Table 5.

Table 5: Representative Age and Savings for Each Age Group (Female/Own) [5]

Age Group	Representative Age	Savings			
		Max	Predicted Value	Predicted Error	Threshold
	A_{rep}	S_{MAX}	S_{th0}	ΔS_{th}	S_{th}
65 - 69	65	506	939.2	-433.2	1403.7
70 - 74	70	1194	729.5	464.5	1194.0
75 - 79	75	800	519.8	280.2	984.3
80 - 84	80	89	310.1	-221.1	774.6
85 -	85	10	100.4	-90.4	564.9

4.4.2 Two-Step Methods to Determine Dissaving Risk

The test data classified in Section 4.3 are utilized to evaluate our proposed method through the following two-step method. On the basis of the value of S_{th} set in the fashion explained in Section 4.4.1, dissaving risk against each test data is evaluated through either (A) primary determination or (B) secondary determination stated as follows:

(A) Primary Determination

The test data whose savings are no less than S_{th0} are determined as having “no dissaving risk.” The results of primary determination are shown in Table 6. Each column indicates as follows:

Table 6: Primary Determination Results [5]

No.	P	Primary Determination Results				
	Actual	Determination	T/F	Result	Abb	Notation
1	$< P_{th}$	No Dissaving Risk	FALSE	Leak Detection 1	LD1	N_{L1}
2	$\geq P_{th}$		TRUE	No Risk 1	NR1	N_{N1}

- “P” provides “$< P_{th}$” or “$\geq P_{th}$.” As stated in Section 4.2.2, P_{th} is the threshold to determine dissaving risk of the test data and is set to 0.1 (0.3, respectively) for Singles (Twos).
- “Determination” provides the result of determining dissaving risk, either “Dissaving Risk” or “No Dissaving Risk.” For Table 7, only “No Dissaving Risk” is provided.
- “T/F” indicates whether determination result is correct or not. If the determination result is correct, this column provides “TRUE,” otherwise it is “FALSE.”
- “Result” is classified as two results; “Leak Detection 1 (LD1)” and “No Risk 1 (NR1).”
- “Abb” indicates the abbreviation of results shown in the column “Result.”
- “Notation” indicates the amount of the data.

(B) Secondary Determination

Secondary determination is processed against the test data that have not been determined in primary determination yet, through the following three methods. The savings of these test data are less than S_{th0} for the respective age group A_{rep}. Estimated SET is named as ESET.

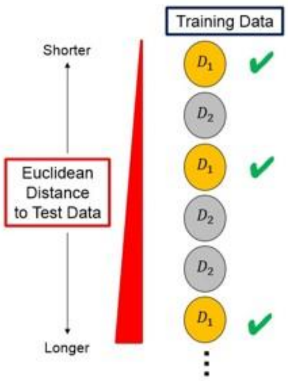
(i) Method 1

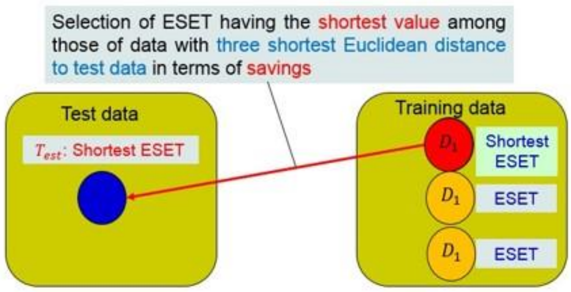
For a test data, *ESET* can be calculated as follows. Suppose that D_{rev} is the Re-dis, I is the amount of income, and S is that of savings, similar to Formula (2), the T_{est} can be calculated by using Formula (9).

$$ESET = \frac{S}{D_{rev} - I} \tag{9}$$

The T_{est} is then chosen out of several *ESETs* of the training data. The procedure of determining T_{est} is illustrated in Figure 2. Firstly, the three training data whose Euclidean distances to one test data are the first to third shortest in terms of savings are extracted from the training data belonging to only D_1 and depicted in Figure 2-(a). The criterion of the training data with first to third shortest are determined through trial and error. Taking the example of Figure 2-(a), training data are sorted in the ascending order of Euclidean distance to the test data. Here, the first (shortest) training data is firstly extracted, as partial set D_1 is assigned to it. The extracted data is shown as an orange circle with a green check mark right next to it. Meanwhile, the second shortest training data is discarded, as partial set D_2 is assigned to it. The disregarded data is illustrated as a gray circle. Similarly, these procedures will be repeated until the extraction of third training data with D_1. In the example of Figure 2-(a), the first, third and sixth shortest training data are extracted and used for the subsequent procedure. *ESET* is then calculated for the three training data.

Among the three *ESETs* extracted, the shortest one is set as T_{est}. In Figure 2-(b), the test data with the shortest is shown as a red circle, while the other two are orange circles. The red arrow then derives from the red circle to a blue circle (the test data) to set the shortest *ESET* as T_{est}.





(a) Extracting Training Data with D_1 (b) Assigning T_{est} as Shortest *ESET*

Figure 2: Illustration of the Procedure of Secondary Determination (Method 1) [6]

T_{est} is then divided by the life expectancy L_{exp} corresponding to the age group and the survey year (1994, 1999, or 2004) of the test data. Provided that P_{est} is the ratio of T_{est} to L_{exp} for the test data, similar to Formula (3), P_{est} is calculated using Formula (9).

$$P_{est} = \frac{T_{est}}{L_{exp}} \tag{9}$$

If P_{est} is less than P_{th}, the threshold to determine dissaving risk, the test data is designated as having "Dissaving risk." Otherwise, it will be regarded as "No dissaving risk." As stated in Section 4.4.2, P_{th} is set to 0.1. is set to 0.1 (0.3, respectively) for Singles (Twos). The results of secondary determination are summarized in Table 7.

The columns are basically the same as provided in Section 4.2.1 and shown in Table 6. Minor differences between Tables 6 and 7 are explained as follows:

- "T/F" provides "TRUE" if the column entitled "Estimated" and "Actual" correspond, otherwise it is "FALSE."
- "Result" is classified as the following four categories: "Correct Detection (CD)," "Incorrect Detection (ID)," "Leak Detection 2 (LD2)" and "No Risk 2 (NR2).

Table 7: Secondary Determination Results [5]

No.	P_{est}	P	Secondary Determination Results				
	Estimated	Actual	Determination	T/F	Result	Abb	Notation
1	$< P_{th}$	$< P_{th}$	Dissaving Risk	TRUE	Correct Detection	CD	N_C
2	$< P_{th}$	$\geq P_{th}$		FALSE	Incorrect Detection	ID	N_I
3	$\geq P_{th}$	$< P_{th}$	No Dissaving Risk	FALSE	Leak Detection 2	LD2	N_{L2}
4	$\geq P_{th}$	$\geq P_{th}$		TRUE	No Risk 2	NR2	N_{N2}

(ii) Method 2 and Method 3

Methods 2 and 3 have both similarities and differences from Method 1. For both methods, for a test data, three data whose distances to the test data are the first to third shortest in terms of savings are extracted from the training data belonging to D_1 or D_2. If two training data belong to D_2 out of the three, the test data is determined as having "no dissaving risk." Otherwise, while the procedures for Methods 2 and 3 are basically similar as that for Method 1, these methods will be processed in a subtly different fashion. For Methods 2 and 3, respectively, the three data whose distances to the test data are the first to third shortest in terms of "savings" and the two-dimensional vector space composed of income and savings are extracted from training data belonging to D_1, similar to Method 1. How to determine T_{est} out of the three ESETs and how to determine dissaving risk of the test data are processed in a similar way as Method 1.

4.4.3 Evaluation of Proposed Method

As for the test data determined as the total six categories shown in Tables 6 and 7, for convenience, NR1 and NR2 (LD1 and LD2, respectively) are collectively named as NR (LD). Here, correct detection rate (CDR) and correct judgment rate (CJR) are calculated as follows.

- CDR means the ratio of the amount of CD to the sum of the test data determined as having dissaving risk in the secondary determination. For the denominator, the data are determined as having "Dissaving Risk" in the column entitled "Determination" in Table 7 matching with the sum of the amount of CD and that of ID. CDR is calculated using Formula (10):

$$CDR = \frac{100 N_C}{N_C + N_I} \tag{10}$$

- CJR indicates the ratio of the sum of the amount of CD and that of NR to the amount of all the analysis data. For the numerator, these data are expressed as "TRUE" in the column entitled "T/F" in Tables 6 and 7. CJR is calculated with Formula (11):

$$CJR = \frac{100(N_C+N_{N1}+N_{N2})}{N_C+N_I+N_{L1}+N_{L2}+N_{N1}+N_{N2}} \tag{11}$$

5 Analysis Results

5.1 Previous Analysis Result and Requirements

In the previous analysis, as an initial step of our proposed method, only the data of single females were tentatively analyzed [6]. With the application of methodology explained in Section 4, the analysis result has conveyed that CJR exceeded 80% for single females.

As explained earlier, merely the data of single females were analyzed. However, the data other than single females must also be used for the further analysis. Thus, in this paper, the data of single males and two-person households are analyzed in a similar fashion as the previous analysis. The details of analysis results for single females will be summarized in subsequent Section 5.2.

5.2 Further Analysis Results

As explained in Section 5.1, the data of single males and two-person households are evaluated in a similar approach as those of single females, using the methodology summarized in Section 4. The analysis results obtained and calculated through Formulas (10) and (11) depicted in Section 4.5 are summarized in Table 8. As stated in Section 5.1, the results of single females are included in Table 8. Comparing the results, as for single males, any of the three method shows approximately similar results for rent, while Method 1 shows the best for own. Meanwhile, as for females, Method 2 (3, respectively) outperforms for rent (own). As for Twos, CDR indicates best with Method 2 for rent, whereas both CDR and CJR show best with Method 3 for rent. As a whole, *CJR* outscores 80% with any methods for Singles. Meanwhile, *CJR* achieves approximately as high as 70% for Twos.

For our future work, the improvement of determination method especially for Twos is required. Since our proposed method has been developed on the basis of ample data that covers elderly people, how to cope without such data would be included in our requirements.

Table 8: Analysis Results

(a) Rent

Methods [%]	Single				Two-person Household	
	Male		Female			
	CDR	CJR	CDR	CJR	CDR	CJR
Method 1	100.0	90.0	56.1	87.8	32.1	59.9
Method 2	100.0	90.0	60.6	89.0	39.5	71.7
Method 3	100.0	90.0	60.0	89.0	33.3	68.4

(b) Own

Methods [%]	Single				Two-person Household	
	Male		Female			
	CDR	CJR	CDR	CJR	CDR	CJR
Method 1	62.5	94.4	30.4	83.5	36.5	68.1
Method 2	50.0	90.7	32.4	86.2	36.2	70.4
Method 3	50.0	90.7	38.6	88.8	37.3	71.5

6 Conclusion

In this paper, we proposed a method to detect and evaluate dissaving risk of elderly people at the age of sixty-five or older through using the anonymous data based on the NSFIE carried out by the MIC in 1994, 1999, and 2004. Here, the analysis data were divided into test data and training data. Test data were then evaluated through a two-step method according to the amount of savings of the test data. In the case of secondary determination, three methods were performed in terms of income and savings. The analysis result conveys that the correct judgment rate exceeds 80% for Singles, while the figure attains mere approximately 70% for two-person households.

For future work, the CDR and CJR must be improved, especially for two-person households. How to deal with insufficient data covering elderly people is also included in our requirements.

Acknowledgement

This research was supported by the Research Institute of Science and Technology for Society within the Japan Science and Technology Agency.

References

[1] H. Ohba, Y. Kadoya, and J. Narumoto, "Influence on the household change Caused by Dementia Decline -Two-year Longitudinal Analysis Using JSTAR Data -" (in Japanese), Proc. of the 20th Japan Society of Geriatric & Gerontological Behavioral Sciences, 2017.

[2] Aomori Prefectural Industrial Technology Research Center, etc., “Dementia Prediction Program and Method” (in Japanese), Japan patent JP2015-138488A, 2015-07-31.

[3] Toshiba Corporation, "Dementia Determination Device, System, Method, and Program" (in Japanese), Japan patent JP2017-104289A, 2017-06-15.

[4] e-Stat: Search for statistics (URL): https://www.e-stat.go.jp/en/stat-search/database?page=1, 2022-8-31.

[5] Yuya Yokoyama and Yasunari Yoshitomi, “Detailed Assessment of Dissaving Risk Against Life Expectancy for Elderly People using Anonymous Data and/or Random Data: A Review,” Advances in Science, Technology and Engineering Systems Journal, Vol.6, No.2, pp.871-886, 2021.

[6] Yuya Yokoyama and Yasunari Yoshitomi, “Using Anonymous Data to Assess Dissaving Risk against Life Expectancy for Elderly People,” Proc. of 10th International Congress on Advanced Applied Informatics (AAI 2021), pp.3-9, 2021.

[7] T. Fujita, S. Ogano, and J. Narumoto, “Dementia and Information” (in Japanese), ISBN978-4-326-44976-7, Keisou Shobou, p.150, 2019.

[8] Sumit Agarwal and Bhashkar Mazumder, “Cognitive Abilities and Household Financial Decision Making,” American Economic Journal: Applied Economics, pp.193-207, 2013.

[9] Dianh Bisdee, Debora Price and Tom Daly, "Coping with Age-related Threats to Role Identity: Older Couples and the Management of Household Money," Journal of Community & Applied Social Psychology, Vol.23, pp.505-518, 2013.

[10] Michael S. Finke, John S. Howe and Sandra J. Huston, "Old Age and the Decline in Financial Literacy," Management Science, Vol.63, No.1, pp. 213-230, 2017.

[11] Joanne W. Hsu and Robert Willies, "Dementia Risk and Financial Decision Making by Older Households: The Impact of Information," Journal of Human Capital, Vol.7, No.4, pp.340-377, 2013.

[12] National Statistics Center (URL), https://www.nstac.go.jp/en/, 2022-8-31.

[13] Life expectancies at specified ages (URL), https://www.mhlw.go.jp/english/database/db-hw/lifetb04/1.html, 2022-8-31.

[14] Consumer Price Index (URL), http://www.stat.go.jp/english/data/cpi/index.html, 2022-8-31.

[15] Yuya Yokoyama and Yasunari Yoshitomi, "Assessment of Dissaving Risk against Life Expectancy for Elderly People through Anonymous Data and Random Data," Proc. of 20th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD2019), pp.274-279, 2019