Panel Data Research Center, Keio University

PDRC Discussion Paper Series

【第６回学生論文コンテストJHPS AWARD 受賞論文：審査員賞】

Analyzing the Impact of Subjective Well-Being on Consumption -Insights from Machine Learning Predictions and Econometric Models-

村中　隼

2025 年 3 月 31 日

DP2024-009
https://www.pdrc.keio.ac.jp/publications/dp/9409/

Panel Data Research Center, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
info@pdrc.keio.ac.jp
15 March, 2025

【第６回学生論文コンテスト JHPS AWARD 受賞論文：審査員賞】
Analyzing the Impact of Subjective Well-Being on Consumption -Insights from Machine Learning Predictions and Econometric Models-

村中 隼
PDRC Keio DP2024-009
2025 年 3 月 31 日
JEL Classification: D12; I31
キーワード: Consumption; Subjective well-being; Machine learning; Dynamic model

【要旨】

Although previous studies have extensively examined how consumption affects subjective well-being (SWB), whether SWB itself influences consumption remains unclear. This study investigates the relationship between SWB and consumption behaviors using the Japan Household Panel Survey (JHPS) and the Keio Household Panel Survey (KHPS). First, this study uses machine learning tools to determine the importance of well-being and socio-economic variables in predicting consumption by ranking the variables based on their importance. Second, based on the variables selected by the machine learning algorithm, econometrics models for panel data are applied to detect causal relationships running from SWB to consumption. Specifically, we apply the fixed effects model and the dynamic panel model to six consumption categories (total, food, food outside, utility, clothing, and leisure) for the overall sample and subsamples by gender and education. The finding from the machine learning models suggests that age, income, wealth, happiness, and preference variables are influential in predicting total consumption, indicating the importance of well-being and preference variables in consumption decisions beyond the effect of income and wealth. The results of the econometric analysis reveal that happiness positively affects total consumption, and that total consumption increases by 5.8% for a one-unit increase in happiness (ranging from 0-10). When we decompose the effect into sub-categories, we find that food outside of home and leisure consumption are driving the positive effects. Moreover, we show that the effect of happiness differs by gender and education: positive for leisure consumption among females and the lower-educated group, positive for food and negative for clothing consumption among males, and no statistically significant effect for the higher-educated group. In further analysis, we examine whether individuals' preferences could be in the pathway from happiness to consumption. Our econometric evidence shows that the effect of happiness on consumption is unaffected by including the preference variables, and that people who discount the future more tend to increase current consumption. Our results suggest that subjective well-being and preference measures should be considered when designing economic policies.

村中 隼
武蔵大学 経済学部

謝辞： 本稿の作成に当たり、慶應義塾大学 経済学部附属経済研究所パネルデータ設計・解析センターから「日本家計パネル調査」（JHPS/KHPS）の個票データを提供して頂いた。

Abstract

Although previous studies have extensively examined how consumption affects subjective well-being (SWB), whether SWB itself influences consumption remains unclear. This study investigates the relationship between SWB and consumption behaviors using the Japan Household Panel Survey (JHPS) and the Keio Household Panel Survey (KHPS). First, this study uses machine learning tools to determine the importance of well-being and socio-economic variables in predicting consumption by ranking the variables based on their importance. Second, based on the variables selected by the machine learning algorithm, econometrics models for panel data are applied to detect causal relationships running from SWB to consumption. Specifically, we apply the fixed effects model and the dynamic panel model to six consumption categories (total, food, food outside, utility, clothing, and leisure) for the overall sample and subsamples by gender and education. The finding from the machine learning models suggests that age, income, wealth, happiness, and preference variables are influential in predicting total consumption, indicating the importance of well-being and preference variables in consumption decisions beyond the effect of income and wealth. The results of the econometric analysis reveal that happiness positively affects total consumption, and that total consumption increases by 5.8% for a one-unit increase in happiness (ranging from 0-10). When we decompose the effect into sub-categories, we find that food outside of home and leisure consumption are driving the positive effects. Moreover, we show that the effect of happiness differs by gender and education: positive for leisure consumption among females and the lower-educated group, positive for food and negative for clothing consumption among males, and no statistically significant effect for the higher-educated group. In further analysis, we examine whether individuals' preferences could be in the pathway from happiness to consumption. Our econometric evidence shows that the effect of happiness on consumption is unaffected by including the preference variables, and that people who discount the future more tend to increase current consumption. Our results suggest that subjective well-being and preference measures should be considered when designing economic policies.

Analyzing the Impact of Subjective Well-Being on Consumption -Insights from Machine Learning Predictions and Econometric Models-

1. Introduction

In traditional economics models, emotional well-being or happiness is typically neglected. However, recent studies in psychology and neuroscience provide strong evidence to suggest that happier people tend to make decisions differently from unhappy people. The recent economic literature that studies the relationship between well-being and business cycles also finds supporting evidence that fear of unemployment and people's perception about future economy can be used to predict unemployment one year later (Blanchflower and Bryson; 2021; 2023). This literature is often dubbed "the economics of the walking about." Furthermore, past studies have explored the effect of happiness on a variety of economic variables. For example, Oswald et al. (2015) conducted experiments in English universities and found that happier individuals exhibited higher productivity. Mohanty (2009) found that workers with positive attitudes directly and indirectly increased their wages in the US. Graham et al. (2004) showed that current happiness positively affects future income and health in Russia.

The main aim of this paper is to examine how subjective well-being or happiness affects consumption behaviors in Japan. Past studies have mostly examined how subjective well-being is affected by consumption (e.g. Wang et al. 2019; DeLeire and Kalil, 2010; Cui, 2018), whereas studies that examine the effect of subjective well-being on consumption are scant (Dominko and Verbic, 2022; Guven, 2012; Goldsmith, 2016). In addition, these studies suffer from several limitations. First, the choice of variables to predict consumption is arbitrary. There are many different variables predicting consumption and there are also many measures of subjective well-being. Existing studies typically select variables without much justification. Second, past studies do not consider the dynamic patterns of consumption that current consumption could be affected by past levels of consumption. Lastly, among the few studies that tackled endogeneity problem of subjective well-being, the method is limited to instrumental variables. The instruments used in past studies include sleep quality and regional sunshine, but these instruments may not be perfect because they could suffer from exogeneity.

We overcome some of the methodological limitations of existing studies by using machine learning methods for variable selection and econometrics analysis for the identification of causal effects. In particular, machine learning methods allow us to systematically rank the relative importance of variables in predicting consumption. However, machine learning methods do not offer an interpretation of the results. By contrast, econometric analysis can be used to address the endogeneity problem and provide a causal interpretation. In particular, we use fixed effects models and a dynamic panel data model to identify the causal relationship between subjective well-being and consumption decisions.

We hypothesize that the effect of subjective well-being influences differs by different types of consumption, such as food, clothing, and entertainment. When people are happier with their lives, they may choose to increase the consumption of certain types of goods (e.g., eat out) and decrease the consumption of other types of goods. Past studies have linked hedonic consumption with subjective well-being (e.g. Zhong and Mitchell, 2012). Another possible channel is that happiness could affect people's time preferences. For example, Ifcher & Zarghamee (2011) conducted an experiment that induced people to have positive affect, and they showed that induced positive affect tends to make people prioritize the future, which contributes to long-run planning and thinking. Other studies showed that time preference can affect consumption patterns. For example, Kossova & Sheluntcova (2024) find that individuals with a high discount rate tend to buy fast food more frequently because they are considered to prioritize short-run satisfaction.

This paper uses the 2011-2022 Japan Household Panel Survey (JHPS) and Keio Household Panel Survey (KHPS) for econometric analysis and machine learning analysis. JHPS and KHPS are longitudinal data that track the same households over time, making it possible to analyze dynamic behaviors of consumption. The panel data offer longitudinal information on detailed consumption categories and wealth-related variables as well as socio-economic and subjective well-being variables. For the machine learning exercise, we rely on 2020 JHPS. We selected the year of 2020 because it is prior to COVID-19, and consumption decisions would not be affected by the pandemic. We included more variables in the machine learning exercise than in the econometrics analysis to sort out the importance of potential covariates.

We offer several contributions to the existing literature. First, methodologically we apply both machine learning and econometrics methods in our empirical analysis. Second, we study the relationship between subjective well-being and consumption in the Japanese context. Lastly, we examine the possible interaction between subjective well-being, gender, education, and detailed categories of consumption, which has not been examined extensively by previous studies.

Examining how subjective well-being fluctuates along the business cycle and how it may, in turn, affect economic activities could provide additional insights for understanding household behaviors during economic downturns or after a public health crisis, such as the COVID-19 pandemic. If emotional well-being affects subsequent consumption and saving decisions, governments should consider this channel when designing policies. As we know, the COVID-19 pandemic has brought large turbulence to public health and caused a major deterioration in mental health and subjective well-being in Europe (Easterlin and O'Connor, 2023；Rossouw and Greyling, 2022; Blanchflower and Bryson, 2022) as well as in Japan (Ishii and Yamamoto, 2024; Sato et al., 2022).[1] The results of this study could be helpful in designing economic policies that reduce the impact of economic downturns on households by incorporating the effect of subjective well-being in their analysis.

2. Relevant Literature

Many studies in the literature examined how consumption affects happiness, but not the other way round. For example, Wang et al. (2019) found that spending on clothes, transportation and communication, and necessities have positive effects on happiness. DeLeire and Kalil (2010) found a positive relationship between leisure expenditure and

---

[1] Easterlin and O'Connor (2023) found a negative association between COVID-19 and life satisfaction in European countries. Rossouw and Greyling (2022) found significant decreases in happiness in the case of the Ukraine war and Covid-19. Blanchflower and Bryson (2022) used the Eurobarometer data to study how people's expectations and life satisfaction change during economic downturns, and they show that the expectation variables appear to be more sensitive to economic fluctuations than life satisfaction. Regarding previous papers using Japanese data, Ishii and Yamamoto (2024) found that the COVID-19 pandemic widened subjective well-being inequality, with low-income groups experiencing declines and the high-income group experiencing improvement. Sato et al. (2022) highlighted that young and high-income young females experienced significantly greater declines in happiness compared to elderly women and low-income young females, respectively. The study also observed that central Japan saw larger declines in happiness compared to southwestern and northeastern regions. These findings underline the economic and social disparities in well-being.

happiness in the U.S. because leisure consumption strengthens social connectedness and mitigates loneliness, leading to increased happiness. It can also be treated as conspicuous consumption that enhances social status. Cui (2018) found that consumption aimed at enhancing social status and strengthening connections with others is associated with greater happiness in China.

The studies that are directly relevant to our topic are summarized below. Dominko and Verbic (2022) examined how subjective well-being affects six consumption categories by using a sample of elderly (50 years and older) from England, and they found that their measure of subjective well-being is positively related to food consumption outside of home and leisure activities, but the categories of consumption related to basic needs (such as rent, utility, or food at home) are not affected. They used quality of sleep as an instrument in a panel data setting. Guven (2012) reached somewhat different conclusions by using a different instrument (regional sunshine) and data from Germany and the Netherlands. While they did not examine consumption directly, they examined the tendency of saving (which is the opposite of consumption), self-control of expenditure, and view about the future. They found that happier people are more likely to save, have a higher marginal propensity to save, have stronger self-control over expenditure, and are more concerned about the future than the present. Their happiness variable was measured on a 1-5 scale, ranging from "very unhappy" to "very happy". Goldsmith (2016) found some weak and positive correlation between happiness and the purchase of non-grocery items using data from the US, but their data are cross-sectional, and their focus is the relationship between five big personalities and happiness instead of overall consumption.

We note that the above studies only offer evidence from Western countries, and we are not aware of any studies using Japanese data. Japan could be an interesting country to examine because the saving rate in Japan had been falling gradually since the 1970s and had started to rise slightly before COVID-19.[2] Understanding saving and consumption behavior in Japan is crucial in understanding the fluctuation of the Japanese economy. By focusing on a non-Western context, our study offers new insights into the global applicability of existing theories on well-being and consumption behavior.

3. Estimation Methods

This study employs a two-step approach to analyze the relationship between happiness and consumption. The first step applies machine learning techniques to determine the importance of the variables in predicting total consumption expenditure. Machine learning methods have the advantage of handling large amounts of data and complex non-linear relationships, but they cannot be used to detect causal relationships among variables. In addition, it has difficulty in providing economically meaningful interpretations. Therefore, in the second step, we utilize econometric models to identify the causal relationship running from happiness to consumption by taking advantage of the panel data structure.

3.1 Machine learning analysis

Machine learning techniques offer unique advantages over traditional econometric approaches. Firstly, machine learning techniques enable us to rank variables by their importance, providing valuable insights into which variables are most influential in explaining consumption behavior. Secondly, cross-validation techniques, which are a common

[2] https://www.jcer.or.jp/english/household-savings-rate-in-japan (Accessed January 5, 2025)

strategy to evaluate and select models based on their performance, can be used to ensure that the selected model is robust to unseen data while preventing the models from overfitting. Thirdly, tree-based methods in machine learning techniques make it possible to capture more sophisticated non-linear relationships. Traditional econometric methods often rely on strong assumptions about the correct functional forms, while tree-based methods do not require specifying functional forms, eliminating the risk of misspecification bias. Lastly, the results from machine learning analysis serve as a foundation for subsequent econometric analysis. In this way, we can avoid selecting variables based on our own judgment.

We attempted four machine learning methods (Lasso, Ridge, Random Forest, and Light GBM) that are often used in applied studies. Firstly, Lasso and Ridge are shrinkage methods based on linear regression, each incorporating a unique penalty term in its algorithm. Secondly, Random Forest is a supervised machine learning method based on a tree-based method with a bagging process. This method builds trees from random subsets of the data and predictors to fit the model, and then aggregates their predictions to improve accuracy (Breiman, 2001). It provides a ranking of variable importance in terms of Mean Decrease in Impurity (Gini impurity). Lastly, Light Gradient Boosting Machine (GBM) is a gradient-boosting framework that constructs trees iteratively and optimizes each tree to correct the errors of its predecessors (Ke et al., 2017). It gives us a ranking of variable importance in terms of Gain that measures the contribution of each variable to the model's predictive power.

To evaluate model performances, we use the cross-validation method by dividing the sample into subsets. Some of them are used to train the model, while others are used for testing the model by calculating the errors. Common cross-validation methods are Leave-One-Out Cross-Validation (LOOCV) and K-Fold Cross-Validation (K-Fold CV). In this study, we apply the K-Fold CV to each machine-learning method, in which samples are divided into k-1 folds for model training and one-fold for model testing. To mitigate the effects of skewed data distributions and enhance the reliability of the results, we apply the 10-Fold CV to the data 10 times. At the same time, we also conduct fine-tuning to identify the optimal parameters for each method, in which parameters minimizing the average Root Mean Squared Error (RMSE) over 10 trials are chosen. These parameters are then used for prediction and to calculate the average RMSE. A lower RMSE means stronger prediction power.

It is necessary to consider the presence of multicollinearity among variables in tree-based methods. As Drobnič et al. (2020) point out, strong multicollinearity makes traditional methods for determining feature importance unreliable. By considering this issue in machine learning, we check the values of the Variance Inflation Factor (VIF) and Spearman's rank correlation heatmap for the happiness and life satisfaction variables. Generally, if VIF is greater than 10, the variables exhibit multicollinearity. We report the results of the machine learning exercise with and without highly correlated variables.

3.2 Econometric analysis

Based on the results of the machine learning exercise, we selected variables that are deemed important in predicting consumption expenditure. By using the selected variables, we conducted econometric analysis to further examine the causal relationship. Since happiness (last week) ranked the highest among all subjective well-being variables, we chose to use it in the analysis below.

To identify the causal relationship between consumption and happiness, we must consider the endogeneity resulting from a reverse causality problem and omitted variable bias. First, consumption is measured for the last month, whereas happiness is measured for the last week, which means it is possible that a higher level of consumption causes happiness to increase (i.e. reverse causality). Second, though this analysis includes a wide range of variables associated with consumption, such as gender, age, wealth, income, health, and region, there could still be unobserved individual characteristics that we could not control, such as optimism and pessimism. These characteristics may be associated with both happiness and consumption. Omitted these variables could cause the estimate to be biased (i.e., omitted variable bias).

Recognizing the potential endogeneity of subjective well-being, previous studies have used the instrumental variables (IV) method (Guven, 2012; Dominko & Verbic, 2022). Specifically, Guven (2012) used regional sunshine as the instrument for happiness, but this study does not employ the same approach. The exogeneity of sunshine as an instrument may be questionable in certain contexts because weather patterns can indirectly affect consumption behaviors through their influence on mood, outdoor activities, and energy use, which could violate instrument validity. Moreover, Dominko and Verbic (2022) used sleep quality as an instrument, assuming that it affects happiness but does not affect consumption through channels other than happiness. As Lemola et al. (2013) pointed out, the variability in sleep duration, rather than the mean total sleep time, is associated with lower levels of subjective well-being. There are two sleep-related variables available in the JHPS: sleep quality and sleep hours. Sleep quality is not available until 2021, which limits its applicability in our sample. Sleep hours is a possible instrument, but theoretically, it can be argued that sleep hours may not be exogenous enough because sleeping time may directly influence the time spent on economic activities, including consumption. Indeed, in our machine learning exercise, we show that the variable of sleep hours is ranked within the top 15 variables in importance. For these reasons, we choose not to conduct the instrumental variable approach as in past studies.

Given the endogeneity problem and the limitations of IVs, this paper uses two alternative methods: a fixed effects (FE) model and a Dynamic Panel Model (DPM). We can control for time-invariant factors because the fixed effects transformation removes the time-constant unobservables from the estimating equation. The DPM considers current consumption as a function of previous consumption and both current and lagged happiness, all of which are treated as endogenous variables in the model.

Specifically, for the FE model, we estimate the following equations,

$$Y_{i,t} = \gamma * Happiness_{i,t} + X_{it}\theta + \alpha_i + \mu_{i,t} \tag{1}$$

where $Y_{i,t}$ represents consumption expenditure for individual *i* in year *t*. Happiness represents happiness last week, in which a larger value means greater happiness. The column vector X_{it} includes age, income in a natural log form, the bad health indicator variable, employed dummy, wealth dummies, region and size of city dummies, and year dummies, and α_i represents individual fixed effects that are potentially correlated with $Happiness_{i,t}$ and X_{it}. The error term $\mu_{i,t}$ is assumed to have zero conditional mean and are uncorrelated with the explanatory variables. γ is our parameter of interest because it measures the effect of happiness on consumption. Because the correlation of α_i and $Happiness_{i,t}$ may not be zero, the estimate of γ may be biased. Therefore, we take the within transformation. The estimating equation after the transformation becomes,

$$Y_{i,t} - \overline{Y_i} = \gamma * (Happiness_{i,t} - \overline{Happiness_i}) + (X_{it} - \overline{X_i})'\theta + \mu_{i,t} \quad (2)$$

In equation (2), the individual fixed effects are subtracted away, and time-invariant unobservables are effectively controlled for. In the FE model, we cannot estimate the effect of time-constant variables, such as gender and education.

For the DPM, we estimate the following equation,

$$Y_{i,t} = \alpha Y_{i,t-1} + \beta_1 Happiness_{i,t} + \beta_2 Happiness_{i,t-1} + \beta_3 Health_{i,t} + \beta_4 Employed_{i,t} + W_{it}\phi + \varepsilon_{it} \quad (3)$$

Then we take the first difference of equation (2), we have,

$$\Delta Y_{i,t} = \alpha \Delta Y_{i,t-1} + \beta_1 \Delta Happiness_{i,t} + \beta_2 \Delta Happiness_{i,t-1} + \beta_3 \Delta Health_{i,t} + \beta_4 \Delta Employed_{i,t} + \Delta W'_{it}\phi + \Delta\varepsilon_{it}. \quad (4)$$

Where $\Delta Y (= Y_t - Y_{t-1})$ indicates the first difference of Y. We note that the lagged dependent variable is included on the right-hand side. In addition to lagged happiness, we also include current happiness. Following the literature (Anderson and Hsiao, 1981; Arellano and Bond, 1991), we use lagged values as instruments for endogenous variables and pre-determined variables (correlated with past errors, but with not current and future errors). For endogenous variables, we instrument them with level variables lagged two periods, and for pre-determined variables, we instrument them with level variables lagged one period. For example, $Y_{i,t-2}$ is used as an instrument for $\Delta Y_{i,t-1}$. $\Delta Happiness_{i,t}$ and $\Delta Health_{i,t}$ and $\Delta Employed_{i,t}$ are treated as endogenous and instrumented with level variables lagged two periods. $Happiness_{i,t-1}$ is treated as pre-determined. The variables included in W_{it} (age, gender, education, income, wealth, region and size of city dummies, and year dummies) are treated as strictly exogenous. β_1 and β_2 are our key parameters of interest because they represent the effect of current and past happiness on consumption. Since the level equation can be added to the difference equation and estimated together as a system (Arellano and Bover, 1995), it allows time-constant variables (such as gender) to be estimated.[3] We assess the validity of the instruments by using the overidentification test, i.e., Sargan test and Hansen test. A large p-value indicates that we fail to reject the null hypothesis that the instruments are valid. The Hansen test is robust to heteroskedasticity in the error term.

[3] In estimation, we applied the two-step estimator with system GMM by using Stata command (xtabond2). We note that the endogenous variables in the level equations are instrumented with lagged differenced variables.

4. Data

4.1 Data description

JHPS and KHPS are longitudinal data that track the same households and individuals over time, making it possible to analyze dynamic behavior. The panel data offer longitudinal information on detailed consumption categories and wealth-related variables as well as socio-economic and subjective well-being variables, enabling causal inference in econometric models. The surveys are conducted in January every year by using a two-stage stratified random sampling method. In the first stage, the whole country is stratified into 24 levels based on regional and city classification, and the sample size is distributed according to the population in each stratum. In the second stage, households are randomly selected from basic resident registers in the chosen survey areas. The KHPS began in 2004 with an initial sample of 4,000 households, and additional cohorts of 1,400 households were added in 2007 and 1,000 in 2012. Following this, the JHPS started in 2009, targeting 4000 households. KHPS focused on occupation, consumption, income, and homeownership, while JHPS emphasized occupations, incomes, education, and health. Since 2014, the two data sets shared the same standardized questions.

4.2 Variables for econometric and machine learning analysis

In JHPS/KHPS, respondents were asked about their total expenditure in the last month and expenditures in multiple categories. In addition to total expenditure, we selected four sub-categories of consumption: food, food outside, utility, clothing. We also constructed another expenditure category and named it leisure, which is the sum of culture, amusement, and entertainment expenditures (stationery, sporting goods, travel, hobbies, allowances, membership fees, other association fees, etc).

Respondents were asked this question, "How your feeling of happiness was during the following periods, on a scale of 0 to 10, with 0 being "having no feeling of happiness at all," and 10 being "having a feeling of complete happiness." The corresponding periods are this week, this year, and whole life. The econometric analysis uses happiness from the last week, as the importance measure from the machine learning exercise shows that happiness for the last week ranks higher than happiness for the last year and life, suggesting that recent happiness is more relevant.

The wealth variable is calculated as total assets minus the total amount of borrowing, where the total asset is the sum of the value of owned land, owned house, savings, stock held in domestic currencies, and securities held in foreign currencies. The total amount of borrowing reflects the money people borrow for various reasons, including possessions of houses, land, durable goods, education, leisure, marriage, self-employment, illness, disaster, living, and other things. The wealth variable follows a very skewed distribution and can be negative if total liability is greater than total asset. We create five dummy variables to indicate the level of wealth, with one dummy variable indicating non-positive wealth, and the other four representing the bottom 25% (wealth level 1) to the top 25% (level 4) of positive wealth.

Other control variables are gender, age, income in the last year measured in ten thousand of yen, dummy for the bad health condition, being employed, region, and size of the city. Income includes all income-related variables (annual employment income, self-employment income, interest/dividends, public pension, etc). Health condition is measured from 1 to 5, in which 1 is the best and 5 is the worst. We convert it into a dummy variable indicating bad health, which includes only 4 or 5. The lower education dummy is defined as a junior high school education or high school diploma, while the higher education dummy is defined as a junior college, technical college degree, a university, or graduate school.

Both consumption and income have been converted to real terms by adjusting for inflation using the Consumer Price Index (CPI).[4] Since consumption is recorded at the household level and happiness is measured for each respondent, the matching would be difficult if the household consists of multiple individuals. To make sure that consumption and happiness are both for the same individual, we limit our focus to single individuals only. This analysis utilizes data spanning from 2011 to 2022, as the happiness variable is not available until 2011.

4.3 Variables for machine learning analysis and further analysis

In addition to the variables included in econometric analysis, machine learning analysis includes more subjective-wellbeing variables, such as happiness for the last year and life, and life satisfaction variables. Regarding life satisfaction, respondents were asked the question, "How you feel about the present situation regarding the following, on a scale of 0 to 10, with 0 "not at all satisfied," and 10 is "fully satisfied." The corresponding situations are household income, employment, housing, amount of leisure time, the way you spend your leisure time, your health, and life overall.

In addition, we also included two preference variables (discount factor and risk preference). Discount factor is measured by this question, "Instead of receiving 10 thousand yen one month later, at least how much would you like to receive 13 months later?". There are eight possible answers, ranging from 9,500 to 14,000 yen. The corresponding annual interest rate is calculated for each category. They are -5%, 0%, 2%, 4%, 6%, 10%, 20%, and 40%, with a higher value representing more discounting the future. In other words, if people value the future more, they would require less compensation in the future, thus a lower discount rate. The risk preference is measured by the following question, "When you go out to a place you have never been to before with your family or friends, what percentage of chance of rain makes you decide to take an umbrella?", with answers ranging from 0-100.

5. Result

5.1 Summary statistics

Table 1 reports summary statistics of the main variables used in our econometrics analysis. Since there are zero expenditures in the subcategories of consumption, the number of observations is smaller compared to the total expenditure. The summary statistics for the machine learning analysis can be found in the Appendix Table A1.

In Table 1, the average values of the expenditures for total, food, food outside, utility, clothing, and leisure in the last month before the interview are ¥243,561, ¥53,024, ¥18,759, ¥25,644, ¥17,806, and ¥40,592, respectively. We note that total expenditure includes more categories than the five subcategories, so the numbers from the subcategories do not add up to the total. The average happiness level is 5.569 out of 10. The average age in our sample is 52 years old, and women constitute 56.9% of the sample. The average annual income was ¥2.9 million, and the average wealth was ¥11.6 million. Furthermore, 18.6% of the respondents faced bad health conditions, 71.1% were employed, 39% were at least college-educated, and 31.6% lived in major cities. 31% of the sample were from the Kanto region.

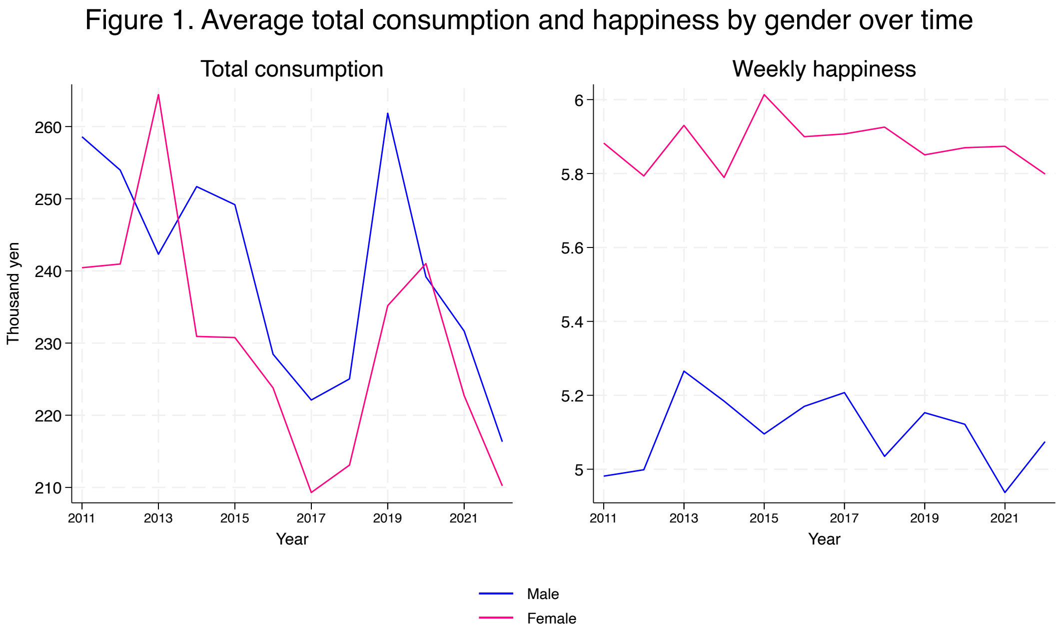
[4] https://tradingeconomics.com/japan/consumer-price-index-cpi (Accessed January 8, 2025)

Table 1. Summary statistics for main variables

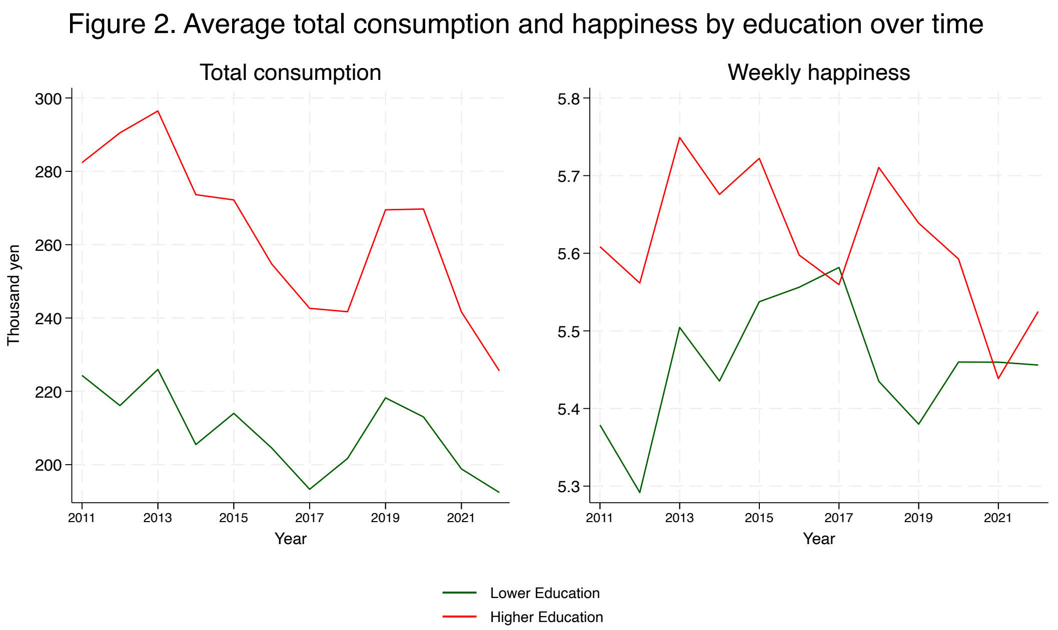
	Observation	Mean	Standard deviation	Minimum	Maximum
Total expenditure (last month, ¥000)	15,339	243.561	246.088	5.030	7171.130
Food expenditure (last month, ¥000)	15,080	53.024	39.956	1.005	1026.612
Food outside expenditure (last month, ¥000)	11,651	18.759	20.208	0.998	322.932
Utility expenditure (last month, ¥000)	14,751	25.644	28.229	1.021	1996.020
Clothing expenditure (last month, ¥000)	9,852	17.806	25.559	0.998	743.750
Leisure expenditure (last month, ¥000)	12,323	40.592	60.332	0.998	2231.250
Happiness (last week)	15,339	5.569	2.417	0	10
Female	15,339	0.569	0.495	0	1
Age	15,339	52.208	17.654	20	96
Income (last year, ¥0000)	15,339	292.300	244.901	0	6645.755
Wealth (¥0000)	15,339	1157.672	5690.384	-44300	512382
Bad health condition	15,339	0.186	0.389	0	1
Being employed	15,339	0.711	0.453	0	1
College educated and above	15,339	0.390	0.488	0	1
Discount factor	12,928	17.32372	15.19577	-5	40
Risk preference	12,928	37.81962	25.42861	0	100
Size of the city (major)	15,339	0.316	0.465	0	1
Size of the city (small)	15,339	0.599	0.490	0	1
Size of the city (towns/villages)	15,339	0.085	0.278	0	1
Region (Hokkaido)	15,339	0.043	0.203	0	1
Region (Tohoku)	15,339	0.073	0.261	0	1
Region (Kanto)	15,339	0.316	0.465	0	1
Region (Chubu)	15,339	0.160	0.367	0	1
Region (Kinki)	15,339	0.180	0.384	0	1
Region (Chugoku)	15,339	0.064	0.245	0	1
Region (Shikoku)	15,339	0.033	0.179	0	1
Region (Kyushu)	15,339	0.131	0.337	0	1

Notes: The data are from the KHPS and JHPS from 2011 to 2022. The data consists of single individuals only.

5.2 Graphs

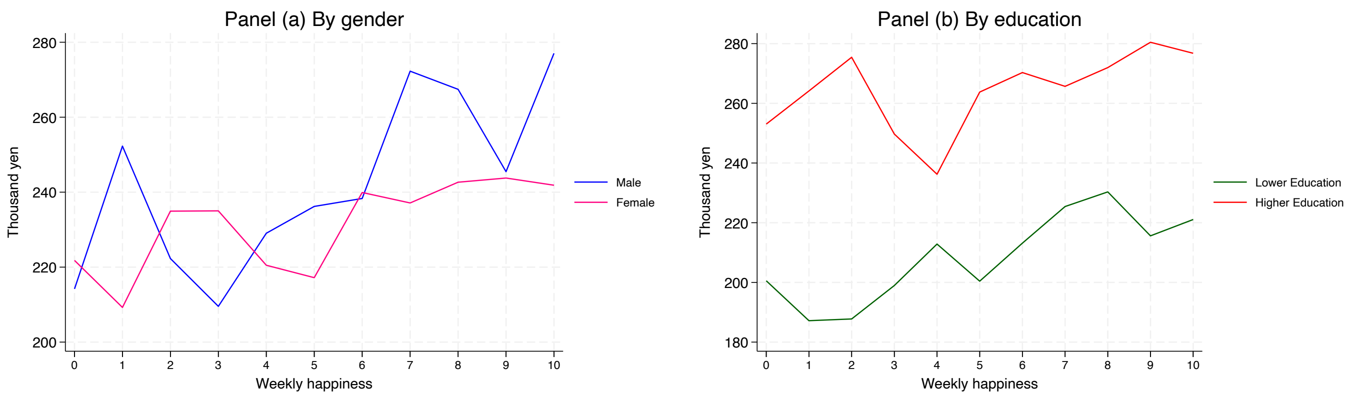


In Figure 1, we plot the time series of total consumption and happiness by gender. In the left panel, we observe that total consumption for both genders followed a similar trend during our sample period except for 2013 and 2020. The average expenditure for males is greater than that for females for most of the years. In addition, both genders experienced notable decline in expenditure from 2021 to 2022, likely due to the COVID-19 pandemic. Turing to the right panel, females reported higher average happiness levels than males in our sample period (5.9 for females and 5.1 for males). In addition, males and females experienced somewhat different trends: females experienced a decline in their happiness from 2021 to 2022, while happiness rebounded in 2022 for males. This could reflect differences in psychological reactions to the pandemic between men and women.



In Figure 2, we plot the time series total consumption and happiness by education level, with lower education indicating high school and less and higher education indicating college or more. In the left panel, we observe that higher-educated individuals exhibit a greater level of consumption than lower-educated individuals. The trend between the two groups is similar though: they both decreased from 2011 to 2017, and rebounded from 2017 to 2019, probably because of anticipatory demand before the consumption tax increase in October 2019 from 8% to 10%; since 2020, total consumption has decreased in both groups. Regarding average happiness (right panel), a similar trend between the two education groups can be seen until 2015, after which higher and lower educated experienced different trends. Additionally, people with higher education experienced higher happiness levels in most years, with the exception of 2017.

Figure 3. Average total consumption for each happiness level



In Figure 3, we present the average consumption for different levels of happiness by gender and education. In Panel (a), total consumption in males increases as their happiness level rises. By contrast, relatively stable total consumption is seen for females across different happiness levels. In Panel (b), there is a clear gap between those with different education levels, and the lower educated group increases their consumption as happiness rises, while the higher educated group experienced fluctuations from 0 to 5 but an upward trend from 6 to 10.

These findings emphasize the importance of considering demographic differences when analyzing the relationship between consumption and happiness.

5.3 Machine learning performances

Table 2. Performances of four machine learning models

Model	Ridge	Lasso	Random Forest	LightGBM
Average RMSE	0.6658	0.6641	0.6642	0.6580

Note: Average RMSE is the RMSE over ten trials.

Table 2 summarizes the results. The average RMSE for lasso and random forest is similar at 0.6641 and 0.6642, respectively, and the lowest average RMSE is recorded for LightGBM. Thus, we use it to select potential determinants for consumption.

5.4 Variable selection

Table 3. VIF for happiness and life satisfaction variables

Variables	VIF
Happiness (last week)	3.2571
Happiness (last year)	4.357
Happiness (life)	2.5807
Life satisfaction (income)	1.7671
Life satisfaction (work)	1.9456
Life satisfaction (housing)	1.8273
Life satisfaction (leisure time)	2.4274
Life satisfaction (leisure quality)	2.8515
Life satisfaction (health)	2.9213
Life satisfaction (overall)	3.9987

Table 3 describes the values of VIF for subjective well-being-related variables included in the machine learning analysis. All variables are below 5, and high VIFs are reported for happiness for the last year and overall life satisfaction, at 4.357 and 3.9987, respectively. Hence, multicollinearity does not appear to be a significant issue for these variables in terms of VIF.

Figure 4. The correlation heatmap for happiness and life satisfaction variables

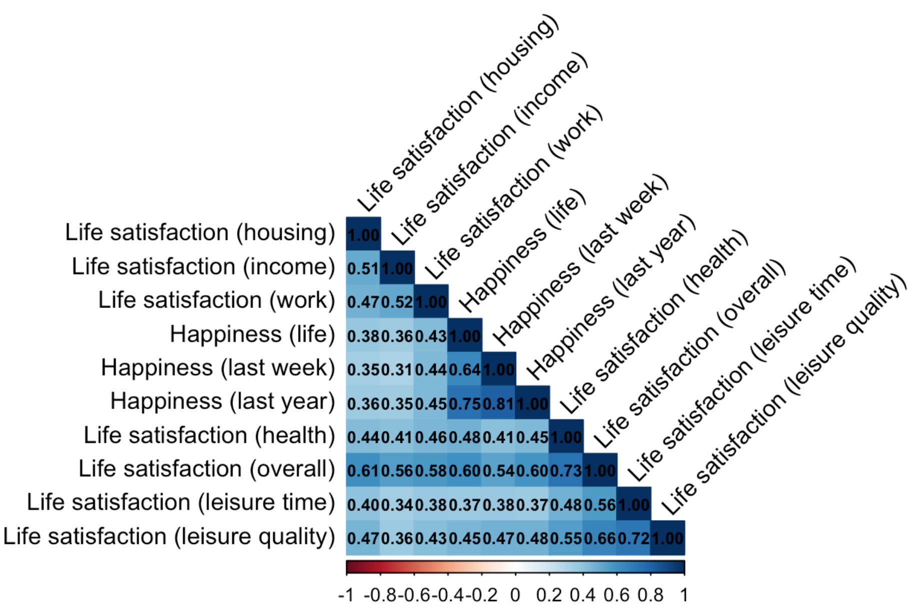
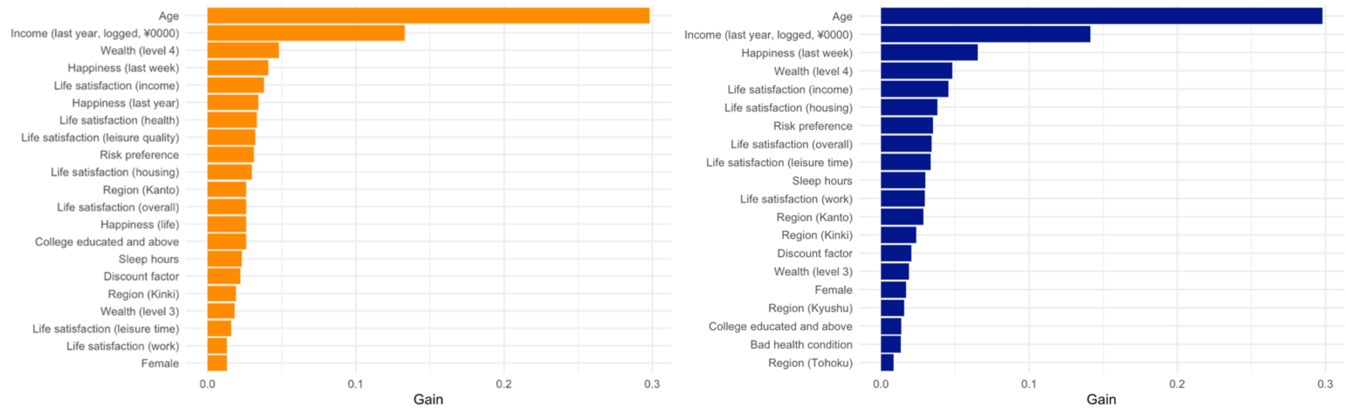


Figure 4 depicts the heatmap in terms of Spearman's rank correlation coefficient. All variables are positively correlated. Happiness variables are highly correlated, and happiness for the last year and the last week recorded the highest value at 0.81. In addition, life satisfaction with overall life and health, and life satisfaction with leisure time and leisure quality, at 0.75, 0.73, and 0.72. Hence, we compute the ranking of important variables with all variables with and without the highly correlated variables (i.e., happiness for the last year and for life, and life satisfaction with leisure quality and with health).

Figure 5. Ranking of variables



Note: we include 35 variables in the left panel, while in the right panel we include the same variables except for happiness for the last year and life, life satisfaction with leisure quality and with health.

In Figure 5, the left panel displays the top 20 most important variables selected and ranked by LightGBM with all variables, including highly correlated variables. We see that age is the most important variable in predicting consumption, followed by income, wealth (level 4), and happiness. Happiness for the last week ranks higher than happiness for the last year and other life satisfaction variables. Other variables in the top 20 include, for example, life satisfaction with income and health, risk preference, region dummy for Kanto, discount factor, education, sleep hours, and gender. In Figure 5, the right panel describes the ranking of important variables without the highly correlated variables, and the ranking is consistent with the left panel. Specifically, age, income, happiness, and wealth are still influential. Rankings for all the variables can be found in Appendix Table A2 and A3.

We note that some of the variables do not show up in traditional economic models in predicting consumption, such as risk preference and the discount factor. We will consider their effects in further analysis. In addition, this machine learning analysis does not imply any causal relationships. In the next section, we will present the results of our econometrics analysis.

5.5 Results for the FE model and DPM

Table 4. Effect of happiness on consumption based on the FE model

	Total	Food	Food outside	Utility	Clothing	Leisure
Happiness	0.004	0.004	0.009**	-0.002	0.002	0.007
	(0.002)	(0.002)	(0.004)	(0.002)	(0.005)	(0.005)
Age	-0.019***	-0.008***	-0.020***	-0.006***	-0.031***	-0.034***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)
Income (Logged)	0.031***	0.012*	0.020**	0.007	0.030***	0.049***
	(0.006)	(0.006)	(0.010)	(0.005)	(0.011)	(0.012)
Wealth (level 1)	0.022	0.035**	0.007	0.013	-0.023	0.031
	(0.015)	(0.016)	(0.023)	(0.014)	(0.029)	(0.030)

Wealth (level 2)	0.026	0.028	-0.023	0.025	0.017	0.044
	(0.018)	(0.018)	(0.027)	(0.017)	(0.034)	(0.034)
Wealth (level 3)	0.060***	0.088***	0.005	0.067***	0.064*	0.103***
	(0.021)	(0.022)	(0.028)	(0.021)	(0.036)	(0.038)
Wealth (level 4)	0.059**	0.107***	0.064**	0.057**	0.088**	0.102**
	(0.024)	(0.025)	(0.031)	(0.026)	(0.039)	(0.042)
Health (bad)	0.027**	0.011	0.049**	0.018	0.022	0.003
	(0.014)	(0.014)	(0.023)	(0.012)	(0.027)	(0.027)
Being employed	0.035*	0.009	0.116***	-0.004	0.069*	0.112***
	(0.020)	(0.021)	(0.031)	(0.018)	(0.036)	(0.039)
N	15339	15080	11651	14751	9852	12323

Note: Robust standard errors are clustered at the individual level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01. This model includes other control variables (dummies for higher education, regions, years, and size of the city).

In Table 4, we present the estimates from the FE model. Happiness is positive and statistically significant at the 5% level only in food outside. Income and the wealth dummies for the top 25% and 50% are statistically significant, with positive coefficients. By contrast, age is negative and statistically significant at the 1% level in all categories of consumption, indicating that older individuals tend to consume less than younger individuals. The bad health dummy is statistically significant in total and food outside consumption with positive coefficients. Being employed is also significant in the food outside and leisure, with positive coefficients, indicating working individuals tend to eat out more and conduct more leisure activities.

Table 5. Effect of happiness on consumption based on the DPM

	Total	Food	Food outside	Utility	Clothing	Leisure
Happiness	0.058**	0.043*	0.075*	0.032	0.023	0.106*
	(0.024)	(0.023)	(0.039)	(0.021)	(0.034)	(0.055)
L.Happiness	-0.006	-0.006	-0.008	-0.006	0.013	0.001
	(0.004)	(0.004)	(0.007)	(0.004)	(0.010)	(0.010)
Age	-0.007***	-0.001	-0.005*	-0.001	-0.012***	-0.008**
	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)
Income (Logged)	0.052**	-0.013	0.049	-0.013	0.085**	0.044
	(0.023)	(0.022)	(0.036)	(0.018)	(0.041)	(0.045)
Wealth (level 1)	0.015	0.052***	-0.013	0.036**	0.026	0.142***
	(0.017)	(0.018)	(0.025)	(0.017)	(0.035)	(0.043)
Wealth (level 2)	0.055*	0.072**	-0.011	0.077***	0.128**	0.160***
	(0.029)	(0.033)	(0.034)	(0.027)	(0.054)	(0.059)
Wealth (level 3)	0.063*	0.092**	-0.004	0.068**	0.175***	0.230***
	(0.032)	(0.036)	(0.036)	(0.033)	(0.051)	(0.059)
Wealth (level 4)	0.118***	0.109***	0.058	0.116***	0.284***	0.277***
	(0.043)	(0.040)	(0.037)	(0.040)	(0.067)	(0.073)

Health (bad)	-0.011	-0.017	-0.142	-0.005	0.190	-0.290
	(0.093)	(0.107)	(0.178)	(0.086)	(0.159)	(0.202)
Being employed	-0.077	0.060	-0.098	0.071	-0.110	-0.039
	(0.139)	(0.135)	(0.227)	(0.110)	(0.267)	(0.262)
L.Consumption	0.462***	0.665***	0.508***	0.667***	0.177	0.206
	(0.114)	(0.101)	(0.105)	(0.113)	(0.134)	(0.126)
N	12342	12012	8257	11685	6135	8765
Sargan test	0.279	0.224	0.715	0.405	0.005	0.010
Hansen test	0.462	0.037	0.820	0.256	0.510	0.242

Note: Robust standard errors are clustered at the individual level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01. This model includes other control variables (gender, dummies for higher education, regional and year dummies, and size of the city).

In Table 5, we present the evidence by using our dynamic panel model (DPM) in which both current and lagged happiness are included. We find that current happiness is statistically significant in total consumption, food, food outside, and leisure consumption, with total consumption at the 5% level and other categories of consumption at the 10% level. For total consumption, an increase in the happiness level by one leads to an increase in the total consumption by 5.8%. By contrast, we do not find any statistical significance for lagged happiness.

For the Hansen test, we do not reject the null hypothesis that the instruments are valid in all cases except for overall food expenditure. For the Sargen test, we reject the null hypothesis only in the cases of clothing and leisure. In addition, the lagged consumptions are statistically significant in total, food, food outside, and utility consumption, which implies the persistency of consumption. Consistent with the FE model, we find a positive effect of income and wealth and a negative effect of age.

By considering the different consumption and happiness patterns in graphs, we further analyze the relationship between consumption and happiness by dividing the sample into sub-groups by gender and education.

Table 6. Effect of happiness on consumption by gender (DPM)

	Total	Food	Food outside	Utility	Clothing	Leisure
Females						
Happiness	0.025	0.022	0.050	0.009	0.045	0.131**
	(0.022)	(0.026)	(0.032)	(0.021)	(0.038)	(0.060)
L.Happiness	-0.002	-0.003	-0.011	-0.005	0.012	-0.001
	(0.005)	(0.005)	(0.009)	(0.005)	(0.010)	(0.013)
L.Consumption	0.584***	0.486***	0.580***	0.605***	0.118	0.307**
	(0.108)	(0.113)	(0.111)	(0.119)	(0.136)	(0.127)
N	7050	6907	4758	6686	3895	5036
Sargan test	0.766	0.078	0.897	0.002	0.010	0.057
Hansen test	0.579	0.030	0.787	0.064	0.419	0.178

Males						
Happiness	0.037	0.055**	0.009	-0.002	-0.091*	-0.012
	(0.026)	(0.025)	(0.044)	(0.021)	(0.047)	(0.052)
L.Happiness	-0.007	-0.011*	0.004	0.001	0.032	0.013
	(0.006)	(0.006)	(0.010)	(0.006)	(0.020)	(0.013)
L.Consumption	0.445***	0.656***	0.436***	0.518***	0.327**	0.086
	(0.111)	(0.092)	(0.128)	(0.143)	(0.140)	(0.135)
N	5292	5105	3499	4999	2240	3729
Sargan test	0.390	0.830	0.503	0.430	0.002	0.015
Hansen test	0.758	0.700	0.723	0.398	0.356	0.618

Note: Robust standard errors are clustered at the individual level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01. This model includes other control variables (age, income (logged), gender, dummies for wealth, the bad health condition, being employed, higher education, regions and year dummies, and size of the city).

In Table 6, we see that the effect of happiness on consumption differs by gender. While current happiness is statistically significant and positive in leisure expenditure for females, it is statistically significant and positive in food consumption for males. We also find a negative effect on clothing expenditure for males, statistically significant at the 10% level. In addition, lagged happiness in males is statistically significant in food at the 10% level with a negative coefficient, indicating that happier people in the last period tend to reduce their current consumption. It is possible that men who experienced lower happiness in the past period increase their food consumption to compensate for the reduced happiness. A large p-value is found for all subsamples for the Hanse test, indicating valid instruments.

Table 7. Effects of happiness on consumption by education (DPM)

	Total	Food	Food outside	Utility	Clothing	Leisure
Lower education						
Happiness	0.047**	0.036	0.054	0.024	0.023	0.136**
	(0.022)	(0.024)	(0.035)	(0.020)	(0.045)	(0.054)
L.Happiness	-0.006	-0.009	-0.003	-0.008*	0.021	0.006
	(0.005)	(0.006)	(0.010)	(0.005)	(0.014)	(0.013)
L.Consumption	0.497***	0.599***	0.285**	0.462***	0.130	0.081
	(0.114)	(0.092)	(0.127)	(0.126)	(0.119)	(0.119)
N	6398	6246	3852	6108	2853	4303
Sargan test	0.672	0.650	0.212	0.362	0.021	0.214
Hansen test	0.876	0.589	0.801	0.603	0.538	0.925
Higher education						
Happiness	0.005	0.014	0.035	-0.010	0.040	0.027
	(0.025)	(0.028)	(0.037)	(0.025)	(0.041)	(0.046)
L.Happiness	0.003	-0.000	-0.003	0.001	0.010	0.008
	(0.006)	(0.006)	(0.010)	(0.006)	(0.013)	(0.014)

L.Consumption	0.381***	0.509***	0.403***	0.598***	0.230*	0.247
	(0.114)	(0.116)	(0.131)	(0.118)	(0.132)	(0.161)
N	4732	4608	3591	4447	2687	3582
Sargan test	0.032	0.119	0.145	0.058	0.306	0.040
Hansen test	0.664	0.065	0.548	0.351	0.913	0.504

Note: Robust standard errors are clustered at the individual level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01. This model includes other control variables (age, income (logged), gender, dummies for wealth, the bad health condition, being employed, region and year dummies, and size of the city).

In Table 7, we present the estimates of the DPM by education. We find that current happiness only has a statistically significant effect on individuals with lower education (positive for total and leisure, negative for utility). By contrast, happiness is not statistically significant in total as well as in subcategories of consumption for the higher educated. Again, the Hansen's test fails to reject the null hypothesis, indicating valid instruments.

6. Further analysis

Based on the machine learning exercise, we find that risk preference and discount factor rank relatively high in the importance measure. They could possibly influence individuals' consumption decisions. For example, highly risk-averse individuals dislike the fluctuation in future income and may save more to prepare for uncertainty in the future. Similarly, individuals with higher discount values discount future consumption more and prioritize current consumption over future consumption, which may make them consume more in the current period. Hence, we include the risk preference and discount factor variables in the DPM in total consumption to analyze consumption behaviors further.

Table 8. Results of the DPM with risk preference and the discount factor by gender and education

	Overall	Females	Males	Lower education	Higher education
Happiness	0.065***	0.036	0.056**	0.062***	0.013
	(0.024)	(0.023)	(0.028)	(0.020)	(0.023)
L.Happiness	-0.007*	-0.003	-0.010	-0.007	0.004
	(0.004)	(0.005)	(0.006)	(0.005)	(0.006)
Discount factor	0.001***	0.001**	0.001*	0.001**	0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Risk preference	0.000	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
L.Consumption	0.474***	0.564***	0.408***	0.525***	0.410***
	(0.122)	(0.109)	(0.119)	(0.114)	(0.111)
N	10909	6203	4706	5605	4226
Sargan test	0.508	0.534	0.683	0.600	0.130
Hansen test	0.693	0.391	0.633	0.872	0.740

Note: Robust standard errors are clustered at the individual level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01. The overall model includes other control variables (age, income (logged), gender, dummies for wealth, the bad health condition, higher education, being employed, regions and year dummies, and size of the city). The model by gender excludes the gender variable, and the model by education excludes the dummy for higher education.

Table 8 presents the estimates of the DPM for the total consumption with risk preference and discount factor by subgroups. We find that the discount factor is positive and statistically significant in all samples, whereas risk preference does not achieve any statistical significance. We note that this result is in contrast with the results from the machine learning exercise in that risk preference is ranked higher than the discount factor. For happiness, statistical significance is achieved in the overall sample, the males, and the lower educated sample, but not the females and higher educated samples. We note that the marginal effect of happiness on total consumption for is larger for males than females, and for lower-educated than higher educated. The coefficient for lagged happiness is mostly negative and statistically insignificant except for the overall sample, which achieves significance at the 10% level.

7. Conclusion & discussion

This study investigated the relationship between consumption and subjective well-being by using Japanese household panel data. While past studies have mostly examined how consumption affects subjective well-being, this paper reverses the question by examining how subjective well-being affects consumption. Facing uncertainty in variable selection, we applied four machine learning methods and chose the method that yielded the best performance to guide us in selecting relevant variables to include in later econometric analysis. By taking advantage of the panel data structure, we were able to tackle the endogeneity problem of subjective well-being by applying the fixed effects model and the dynamic panel model. We also considered the heterogeneous effects of happiness on consumption by gender and education separately.

In machine learning analysis, age is the dominating factor in predicting consumption patterns, followed by income and wealth. Happiness for the last week ranks the fourth among all variables included, which is also the highest among all subjective well-being-related variables. Interestingly, preference-related variables (risk preference and discount factors) achieve higher ranks than many socio-economic variables, such as, education and employment. The high importance of happiness for the last week may imply that short-term emotional well-being can have a meaningful impact on consumption decisions. Preference-related variables as relatively important variables may indicate that psychological factors and behavioral preferences could influence consumption beyond traditional socio-economic variables.

In econometric analysis, the results of the FE model and DPM consistently show that current happiness increases food consumed outside and leisure consumption (such as, entertainment). This finding is consistent with the findings of Dominko and Verbic (2022), though their sample consists of elderly living in the U.K.. Our estimates indicate that an increase in the happiness level by one leads to an increase in total consumption by 5.8%. We also find that current happiness affects consumption differently across genders and education levels. Notably, leisure consumption is positively affected by current happiness for females but not for males. For males, current happiness increases food consumption and reduces clothing consumption. One possible reason could be the differences in consumption preferences and social relationships. For example, women may prioritize leisure activities that involve social interactions, such as spending time with friends and relatives. Furthermore, the insignificance of happiness among people with higher education could be attributed to consumption smoothness, indicating that higher-educated individuals are less likely to be sensitive to emotional fluctuations.

In further analysis, we include preference-related variables in the dynamic panel model. The effect of happiness on total consumption holds in these models as well. We find that the discount factor achieved statistical significance in the overall sample except for the higher-educated group. The positive coefficient suggests that how much people discount the future could affect current assumption. Different from the results of the machine learning exercise, we could not find the risk preference variable to be statistically significant in predicting consumption.

There are several limitations. First, our sample only consists of singles because including married couples will necessarily make the matching between subjective well-being and consumption difficult. Therefore, we choose not to include married couples in this paper. Second, the JHPS is conducted every year and both subjective well-being and consumption are self-reported. The frequency of the JHPS data and the measurement errors could make it difficult to identify day-to-day emotional fluctuations and its impacts on consumption expenditure. Future studies could explore higher frequency data to better capture the short-term fluctuations in subjective well-being.

References

Anderson, T. W., & Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*, *76*(375), 598–606.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277–297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*(1), 29–51.

Blanchflower, D. G., & Bryson, A. (2021). The economics of walking about and predicting unemployment. Working Paper 29172 http://www.nber.org/papers/w29172

Blanchflower, D. G., & Bryson, A. (2022). The economics of walking about and predicting unemployment in the USA. *National Institute Economic Review*, 1–24.

Blanchflower, D. G., & Bryson, A. (2023). Wellbeing, Expectations and Unemployment in Europe. Working Paper 32006, http://www.nber.org/papers/w32006

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32.

Cui, Z. (2018). Happiness and consumption: Evidence from China. *International Review of Economics*, *65*(4), 403–419.

DeLeire, T., & Kalil, A. (2010). Does consumption buy happiness? Evidence from the United States. *International Review of Economics*, *57*, 163–176.

Dominko, M., & Verbič, M. (2022). The effect of subjective well-being on consumption behavior. *Journal of Consumer Affairs*, *56*(2), 876-898.

Drobnič, F., Kos, A., & Pustišek, M. (2020). On the interpretability of machine learning models and experimental feature selection in case of multicollinear data. *Electronics*, *9*(5), 761.

Easterlin, R. A., & O'Connor, K. J. (2023). Three years of COVID-19 and life satisfaction in Europe: A macro view. *Proceedings of the National Academy of Sciences*, *120*(19), e2300717120.

Goldsmith, R. (2016). The big five, happiness, and shopping. *Journal of Retailing and Consumer Services*, 31, 52–61.

Graham, C., Eggers, A., & Sukhtankar, S. (2004). Does happiness pay?: An exploration based on panel data from Russia. *Journal of Economic Behavior & Organization*, *55*(3), 319-342.

Guven, C. (2012). Reversing the question: Does happiness affect consumption and savings behavior? *Journal of Economic Psychology*, *33*(4), 701–717.

Ifcher, J., & Zarghamee, H. (2011). Happiness and time preference: The effect of positive affect in a random-assignment experiment. *American Economic Review*, *101*(7), 3109–3129.

Ishii, K., & Yamamoto, I. (2024). Trends in Income and Well-Being Inequality During the COVID-19 Pandemic in Japan. *Social Indicators Research*, 1–25.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 3149-3157).

Kossova, T., & Sheluntcova, M. (2024). Fast-food consumption and individual time preferences in Russia: evidence for the social policy. *Journal of Economic Studies*, *51*(5), 965–978.

Lemola, S., Ledermann, T., & Friedman, E. M. (2013). Variability of sleep duration is related to subjective sleep quality and subjective well-being: an actigraphy study. *PloS one*, *8*(8), e71292.

Mohanty, M. S. (2009). Effects of positive attitude on happiness and wage: Evidence from the US data. *Journal of Economic Psychology*, *30*(6), 884–897.

Oswald, A. J., Proto, E., & Sgroi, D. (2015). Happiness and productivity. *Journal of Labor Economics*, *33*(4), 789–822.

Rossouw, S., & Greyling, T. (2022). *Collective emotions and macro-level shocks: COVID-19 vs the Ukrainian war* (No. 1210) [Discussion paper]. Global Labor Organization (GLO). https://www.econstor.eu/handle/10419/267148

Sato, T., Li, A., & Matsuda, Y. (2022). Space Time Analysis of Subjective Well-being in the COVID-19 Outbreak in Japan. *Interdisciplinary Information Sciences*, *28*(1), 25–33.

Wang, H., Cheng, Z., & Smyth, R. (2019). Consumption and happiness. *The Journal of Development Studies*, *55* (1), 120–136.

Zhong, J.Y. & Mitchell, V.W. (2012) Does consumer well-being affect hedonic consumption? *Psychology & Marketing,* 29(8), 583–594.

Appendix

Table A1. Summary statistics for machine learning

	Mean	Standard deviation	Minimum	Maximum
Total expenditure (last month, ¥000)	245.356	233.707	23	3746
Income (last year, ¥0000)	293.954	233.879	0	3876
Female	0.559	0.497	0	1
Age	50.471	18.290	21	92
Wealth (non positive)	0.441	0.497	0	1
Wealth (level 1)	0.197	0.398	0	1
Wealth (level 2)	0.100	0.300	0	1
Wealth (level 3)	0.111	0.315	0	1
Wealth (level 4)	0.151	0.358	0	1
Being employed	0.763	0.426	0	1
College educated and above	0.443	0.497	0	1
Happiness (last week)	5.533	2.421	0	10
Happiness (last year)	5.649	2.300	0	10
Happiness (life)	5.876	2.093	0	10
Life satisfaction (income)	4.523	2.649	0	10
Life satisfaction (work)	5.005	2.630	0	10
Life satisfaction (housing)	5.980	2.520	0	10
Life satisfaction (leisure time)	5.625	2.482	0	10
Life satisfaction (leisure quality)	5.663	2.388	0	10
Life satisfaction (health)	5.703	2.487	0	10
Life satisfaction (overall)	5.809	2.261	0	10
Discount factor	17.528	15.021	-5	40
Risk preference	37.626	25.623	0	100
Sleep hours	6.492	1.173	2	12
Bad health condition	0.170	0.376	0	1
Size of the city (major)	0.328	0.470	0	1
Size of the city (small)	0.589	0.492	0	1
Size of the city (towns/villages)	0.084	0.277	0	1
Region (Hokkaido)	0.045	0.208	0	1
Region (Tohoku)	0.064	0.245	0	1
Region (Kanto)	0.320	0.467	0	1
Region (Chubu)	0.166	0.373	0	1
Region (Kinki)	0.192	0.394	0	1
Region (Chugoku)	0.058	0.234	0	1

Region (Shikoku)	0.030	0.171	0	1
Region (Kyushu)	0.124	0.330	0	1

Table A2. The ranking of variable importance

LightGBM	Gain
Age	0.298
Income (last year, logged, ¥0000)	0.133
Wealth (level 4)	0.048
Happiness (last week)	0.041
Life satisfaction (income)	0.038
Happiness (last year)	0.034
Life satisfaction (health)	0.033
Life satisfaction (leisure quality)	0.032
Risk preference	0.031
Life satisfaction (housing)	0.03
Life satisfaction (overall)	0.026
College educated and above	0.026
Happiness (life)	0.026
Region (Kanto)	0.026
Sleep hours	0.023
Discount factor	0.022
Region (Kinki)	0.019
Wealth (level 3)	0.018
Life satisfaction (leisure time)	0.016
Female	0.013
Life satisfaction (work)	0.013

Table A3. The ranking of variable importance after dropping highly correlated variables

LightGBM	Gain
Age	0.2978
Income (last year, logged, ¥0000)	0.1412
Happiness (last week)	0.0654
Wealth (level 4)	0.0481
Life satisfaction (income)	0.0455
Life satisfaction (housing)	0.0381

Risk preference	0.035
Life satisfaction (overall)	0.0342
Life satisfaction (leisure time)	0.0334
Sleep hours	0.0299
Life satisfaction (work)	0.0295
Region (Kanto)	0.0284
Region (Kinki)	0.0238
Discount factor	0.0205
Wealth (level 3)	0.0187
Female	0.0167
Region (Kyushu)	0.0157
College educated and above	0.0137
Bad health condition	0.0132
Region (Tohoku)	0.0083