Exploring Income and Class Effects on Happiness

Qifei Yuan

QMSS GR5015 Data Analysis for the Social Sciences

Professor Gregory Eirich

December 20, 2021

Table of Contents

1. Introduction

1.1 Research Topic and Question

“Can money buy happiness?” Such a classic debate has received considerable attention from both academic scholars and the general public over time. The relationship between material wellbeing and subjective wellbeing has been a complicated social phenomenon. Researchers have developed different opinions on the income effects on happiness. A popular theory argues that happiness rises with income until reaching a certain threshold. (Kahneman & Deaton, 2010) However, others have found opposite evidence and concluded a constant positive connection between income and happiness. In other words, they believe that money can keep buying happiness without any upper limit.

While it is commonly believed that affluent people are happier, such income effects on happiness cannot be isolated from other non-monetary factors. Income can be viewed as an outcome of social and cultural capitals, such as educational attainment and family occupational background. Family income also differs from individual income. Compared with family income, individual income experiences more temporary and longer-term fluctuations. As individuals share aggregate resources among family members, family income is a more appropriate representation of a person’s economic situation.

In addition to income, social class is another phenomenal indicator of socioeconomic status. The social class combines both objective material substances and subjective values people attached to themselves. The concept of the social class comprises several socio-economic aspects. Class includes material resources (income and wealth), access to social capital (networks and opportunities), and social meaning (moral values, lifestyles, preferences). (Mistry et al., 2021)

This study uses General Social Survey (GSS) in 2018, comprising 1419 respondents in the sample. The survey measures people’s general happiness in the relatively long term. Individuals report their subjective well-being by choosing among “very happy”, “pretty happy”, and “not too happy”.

In conclusion, while family income is an objective measure of one's financial resources, self-identified class is a more subjective perception of social position. This study aims to examine how individuals' long-term happiness is related to (1) objective family income level and (2) subjective social classes. I will also compare the happiness mechanism among income, class, and other socioeconomic factors such as educational degree and age.

1.3 Hypothesis

1. Individuals living in higher-income families are generally happier than those in lower-income families.

2. Individuals who perceive themselves to be in higher social classes are generally happier than those who self-identified as lower classes.

2. Description of Data Set and Variables

2.1 Dependent Variable Encoding

happy is the main dependent variable for both regressions. It measures an individual's self-reported happiness degree in the relatively long term. The question asks "taken all together, how would you say things are these days? Would you say that you are very happy, pretty happy or not too happy?" The GSS survey contains several questions relating to individuals' psychological well-being. Some questions focus on satisfaction levels in particular aspects such as work and financial situations. Other questions have specific time frames such as "in the past seven days" and emphasize the short-term emotional states. However, this research aims to examine the long-term effects of income and social class on one's overall happiness. Compared to other similar variables, *happy* is phrased without specific reference to income or class level, which allows one to explore the underlying associations between overall happiness and other factors.
happy is recoded by excluding "don't know" and missing answers. It is converted from character answers to a numerical scale. The magnitude of the number increases with happiness levels. Specifically, 1= not too happy; 2 = pretty happy; 3= very happy.

2.2 Key Independent Variables

(1) *ln_income:*

ln_income is a numerical variable transformed from *income16,* which asks "in which of these groups did your total family income, from all sources, fall last year before taxes, that is?" *income16* captures the respondent's family annual income before taxes or other deductions. Original answers of *income16* contain 26 income brackets, with the lowest group as "below $1000" and the highest group as "$170,000 or over". To develop an appropriate income measurement used in regression, I performed a two-step transformation for this variable. First, I assign the midpoint dollar value of the category to respondents, which is a standard research practice. Using midpoints as the measure of central tendency in income categories has been justified in GSS methodological report No. 64. (Ligon, 1989) For instance, I designate the $500 for the lowest category, which covers annual income below $1000. For the highest category, I assign an income 1.5 times the lower limit of that category, which is a common imputation method in GSS research. (Firebaugh G & Tach L., 2012) In other words, the income for individuals in the highest category is coded as $255,000 ($1.5*$170,000).

The second step is taking the natural logarithm of the midpoint income value. Log transformation could efficiently normalize the income distribution, rendering more appropriate variables in regression models. Compared to the raw midpoint income, *ln_income* is closer to an ideal normal distribution (skewness = -1.2 and kurtosis = 2.57). While imputing income from categories may result in biased estimation of the true income effect, such a problem is diminished by logged income. (Firebaugh G & Tach L., 2012) Overall, I expect a significant positive relationship between *ln_income* and happiness level at a 95% confidence level.

(2) *class:*

The variable *class* measures an individual's self-perceived social status. The question asks "If you were asked to use one of four names for your social class, which would you say you belong in: The lower class, the working class, the middle class or the upper class?" *Class* is recoded from character answers to a numerical scale. A higher number signifies higher social class. (1= "lower class"; 2= "working class"; 3= "middle class"; 4= "upper class") I expect people from higher social classes to be significantly happier than those from lower classes, with a 95% confidence level.

2.4 Controlled Variables

(1) ***female*****:** respondent's gender. (1= female; 0= male) I expect females to be happier than males. My expectation derives from traditional research that reported a higher level of happiness for females. (Stevenson & Wolfers, 2008)

(2) ***educ*****:** respondent's education level. The variable *educ* measures an individual's highest degree of education, by asking the number of formal schooling years completed. Responses of *educ* range from 0 – 20, indicating from "no formal school" to "8 years of college". The answer of "do not know" has been excluded. I expect education to be positively associated with the happiness level.

(3) ***white*****:** respondent's race. (1= white; 0 = black and other)

(4) ***good_health*****:** respondent's self-perceived health status. (1= excellent and good; 2= fair and poor) This is recoded from the *health* variable, which asks individuals to assess their own health status and choose on four scales from excellent to poor. Answers of "not applicable" and "don't know" are excluded. I expect those in good health are more likely to be happy than those suffering from diseases. Individuals who bear from poor health conditions may incur additional costs or concerns in their lives. High higher medical costs and emotional burdens could possibly lead to more frequent feelings of depression.

(5) ***age*****:** respondent's age. With a mean of 49.1 years old, the *age* ranges from 18 to 89 years old.

(6) ***employed*****:** respondent's current employment status. (1= currently employed, both full-time and part-time; 0= currently unemployed) I recoded the GSS original *wrkstat* variable by reconstructing categories. Instead of using the previous 8 categories, I decide on 2 categories: "employed" and "unemployed". The new category "employed" contains both "working full time" and "working part time." The new "unemployed" category comprises all of the rest situations. Specifically, a person is deemed unemployed if he/she is "with a job but not at work because of temporary illness, vacation, strike", "unemployed, laid off, looking for work", "retired", "in school", "keeping house", and "other." The category of "no answer" is excluded from the model. I expect the "unemployed" individuals to be less likely to feel happy than

“employed” and “other”. Such an expectation is supported by Firebaugh G & Tach L, who conclude that “the unemployed are notably unhappier.”

(7) ***relig***: respondent’s religious status. (1= religious; 0= non-religious) I recoded the GSS variable *relig* into a binary variable. The original *relig* focuses on the different religious preferences while I am only interested in whether one holds the religious belief or not. Therefore, the “none” in original *relig* are directly extracted as “non-religious” (*relig* =0), and all other choices (Catholic, Jewish, etc.) are combined together as “*relig* = 1”. The general expectation from literature is that religious people are happier than nonreligious ones. (Hout & Greeley, 2012) Religious attachment brings people the meaning of lives and a sense of belonging, which could be sources of well-being and happiness feelings.

3. Descriptive Statistics

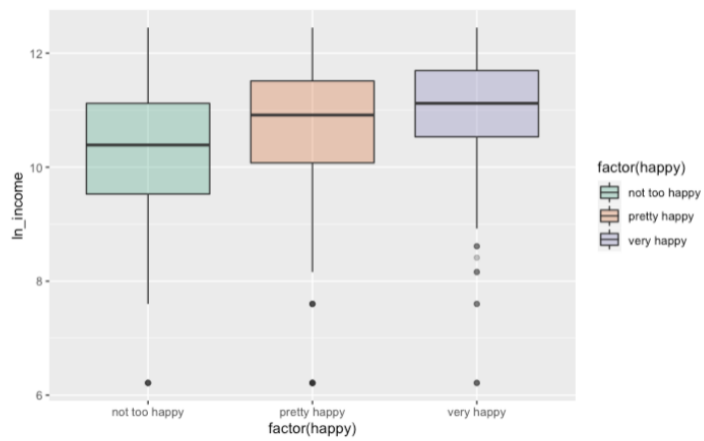
Table 3 presents the descriptive statistics, including mean, standard deviation, median, minimum, maximum, range, skewness, and kurtosis of variables.

The dependent variable, *happy* has a mean of 2.15, indicating that on average people feel a little bit more than “pretty happy”. According to the survey result, 54.3% of all respondents report middle-level happiness, and 30.4% report feeling “very happy”. The distribution of *happy* is approximately symmetric, with skewness = -0.17 $\in$ (-0.5, 0.5). Compared with normal distribution, *happy* is less peaked distributed with shorter and lighter tails (kurtosis = -0.74 < 0) Standard deviation is 0.66, meaning the typical distance between each data point and the mean is 0.66.

The first key independent variable, *ln_income* has a mean of 10.75 and a standard deviation of 1.14. *ln_income* is negatively skewed (skewness = -1.2 < -0.5), indicating a long tail on the left. The fact that the mean is smaller than the median also proves the occurrence of negative outliers. In addition, kurtosis is 2.57 (> 0) and therefore the distribution is more peaked than normal, followed by heavy tails. Notably, while *ln_income* is not an ideal normal distribution, it is more symmetric than the numerical midpoint income distribution (skewness = -1.48 < -1.2). Besides, by shrinking the variable range, the log transformation of income also reduces the effects of extreme outliers in driving the model result.

According to Figure 1, the average *ln_income* increases from “not too happy” to “very happy.” The survey answers show that “very happy” individuals have a higher average income than the less happy two groups, and “not too happy” individuals have the lowest average income. The interquartile range is the smallest in the “very happy” group, indicating a concentration in the high-income family but with a few outliers from the low-income family.

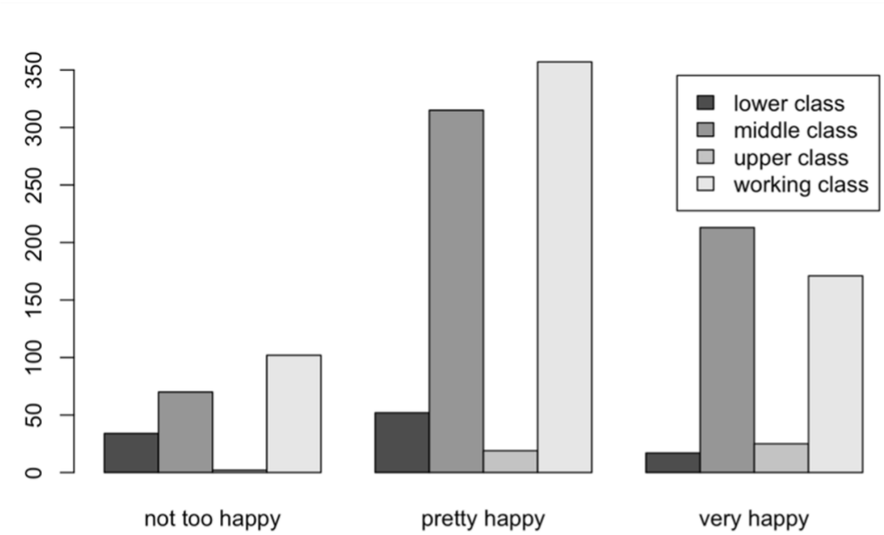
Figure 1. *happy* vs. *ln_income* distribution



The second key independent *class* has a mean of 2.42, which is between “working class” and “middle class”. In fact, 45.24% of respondents perceive themselves to be “working class” and 43.6% identify as “middle class”. For the other two categories, “lower class” has 7.82% response, and “upper class” has 3.31% response. The skewness of -0.15 indicates a relatively symmetric distribution and the kurtosis of -0.3 indicates flatter distribution than normal.

Figure 2 presents the survey response distribution of *happy* and *class*. In the “pretty happy” group, the working class and middle class take the majority. Interestingly, the number of working class is greater than the middle class in the “pretty happy” group, but such ranking is flipped in the “very happy” group. By comparison, the upper class is almost absent in “not too happy”, and primarily present in “very happy.” By distribution of the survey sample, one can notice a positive relationship between class and happiness.

Figure 2. *happy* vs. *class* distribution



For controlled variables, the average age of all samples is 48.15 years old, ranging from 18 to 89 years old. This indicates that all respondents are adults and the majority are middle-aged. Education years range from 0 to 20 years, with a standard deviation of 2.89 years. On average, respondents completed 13.88 years of formal schooling, which is equivalent to slightly higher than a high school degree. In addition, more respondents are religious, employed, whites, and in good health conditions. There are slightly more females (55.25%) than males in the samples.

Table 3. Descriptive Statistics Table for Variables

Descriptive Statistics 1

Statistics	N	Mean	St. Dev	Median	Min	Max	Range	Skew	Kurtosis
happy	1419	2.15	0.66	2	1	3	2	-0.17	-0.74
ln_income	1419	10.75	1.14	10.92	6.21	12.45	6.23	-1.2	2.57
class	1419	2.42	0.68	2	1	4	3	-0.15	-0.3
age	1419	48.15	17.96	46	18	89	71	0.31	-0.88
educ	1419	13.88	2.89	14	0	20	20	-0.36	1.5
relig	1419	0.75	0.43	1	0	1	1	-1.17	-0.62
female	1419	0.55	0.5	1	0	1	1	-0.21	-1.96
employed	1419	0.61	0.49	1	0	1	1	-0.44	-1.81
good_health	1419	0.72	0.45	1	0	1	1	-0.99	-1.02
white	1419	0.73	0.44	1	0	1	1	-1.05	-0.89

4. Initial Models

4.1 Model introduction: Multiple Linear Regression Models

I performed two multiple linear regression models as a starting point. For these two models, the dependent variable *happy* is defined on a numerical scale, ranging from 1 to 4. For initial data exploration, categorical variables are converted to dummy variables. If the variable originally contains multiple categories, it is regrouped into two opposing categories, and used as a binary variable in these first two models. By reducing the number of categories into two major groups, I develop a simple representation of variables, and thus better discover any prominent effects.

While the two models have the same dependent variable, they differ in key independent variables. The first model uses family income while the second model examines social class. All other control variables (socioeconomic indicators) are the same in these two models.

Model 1: Association between Happiness Level and Income

The coefficients are significant for *ln_income* and *good_health* at 99% confidence level, for *employed* at 95% confidence level, and for *relig* at 90% confidence level. All other coefficients on controlled variables are statistically insignificant at 95% confidence level, including *educ*, *age*, *female*, and *white*. The adjusted R-square for model 1 is 0.091 In other words, 9.1% of the variance in the dependent variable *happy* is explained by all independent variables in the model.

As table 4 shows, the coefficient of *ln_income* is 0.122. This means that a 1% increase in income is associated with 0.00122 (0.122/100) higher points in happiness on average, with all else constant. Such a positive relationship matches my expectation, but the income effect is minimal here.

The directions for most variables confirm my expectation, except for *employed*. Specifically, one more schooling year leads to an average of 0.002 increase in the happiness level, with all else constant. One year older in age is associated with a 0.0003 decrease in happiness level on average, controlling for other variables. Compared to males, females have 0.019 points higher in happiness on average, holding others constant. Whites have a 0.028 higher happiness point than non-white (blacks and others) on average, ceteris paribus. Those with good or excellent health conditions are 0.263 points happier than those with fair or poor health states, on average, controlling for all other variables. Individuals with religious beliefs are on average 0.079 points happier than those non-religious ones, all else constant.

However, the negative coefficient of *employed* does not correspond with my previous beliefs. The model implies that being employed makes individuals 0.084 points less happy compared to the unemployed on average, controlling for all other variables. Such coefficient is significant at 95% confidence level, which further confounds me and pushes me to find a better explanation.

Model 2: Association between Happiness Level and Class

The coefficients for *class* and *good_health* are significant at 99% confidence level. It is noteworthy *educ* becomes significant in this second model, at 90% confidence level. In addition, *relig* increases its confidence level to 95% in model 2, while *employed* becomes an insignificant predictor of happiness in this model. R-square for model 2 is slightly lower than model 1. The adjusted R-square of 0.070 represents that 7% of the variance in the dependent variable is explained by the model.

Controlling for all other variables, one higher class is associated with 0.117 points more happiness on average. One more year of formal education increases people's happiness by 0.01 on average, holding all else constant. As one becomes older by 1 year, one tends to become less happy by 0.0003 points on average, all else equal. On average, females are 0.016 points happier than males, controlling for other variables. Compared to non-whites (blacks and others), whites are 0.059 points higher in happiness scale on average, all else constant. Being employed decreases one's happiness by an average of 0.015 points, ceteris paribus. People who perceive themselves as having good and excellent health are on average 0.275 points happier than people who identify as fair or poor health, ceteris paribus. Religious people are on average happier than non-religious people by 0.09 points on the scale, on average, controlling for all other variables.

Table 4. Summary of Initial Models

Initial Linear Models

	Dependent variable:	
	happy	
	(1)	(2)
ln_income	0.122***	
	(0.017)	
class		0.117***
		(0.027)
educ	0.002	0.010*
	(0.006)	(0.006)
age	-0.0003	-0.0003
	(0.001)	(0.001)
female	0.019	0.016
	(0.034)	(0.034)
white	0.028	0.059
	(0.039)	(0.039)
employed	-0.084**	-0.015
	(0.038)	(0.038)
good_health	0.263***	0.275***
	(0.039)	(0.039)
relig	0.079*	0.090**
	(0.040)	(0.040)
Constant	0.600***	1.430***
	(0.168)	(0.109)
Observations	1,419	1,419
R^2	0.096	0.075
Adjusted R^2	0.091	0.070
Residual Std. Error (df = 1410)	0.628	0.635
F Statistic (df = 8; 1410)	18.736***	14.278***
Note:	*p<0.1; **p<0.05; ***p<0.01	

4.2 Result explanation

There are some interesting findings after comparing the two models.

First, the signs of coefficients for each variable remain consistent across both models, indicating a general underlying trend of relationships. The directions of impacts for all these variable do not change based on whether income or class is incorporated in the model.

Second, the magnitudes of *educ* and *employed* have relatively large difference across two models. The coefficients of *educ* are 0.002 and 0.010 respectively. And the coefficients of *employed* are -0.084 and -0.015 in the two models. The large difference in effects signifies potential underlying relationships between these two controlled variables with different key independent variables.

However, this model is insufficient for a few reasons. First, I risked assuming linear relationships between independent variables (*ln_income* and *class*) and happiness level, whilst ignoring other modeling possibilities. Second, the other assumptions for OLS models are not tested. For instance, initial models assume homoscedasticity and normally distributed error without further validation on these assumptions. Third, the interpretation of coefficients is confusing. The linear regression model can only evaluate effects by “the magnitude of points more or less happy”. However, the dependent variable *happy* is categorical. Appropriate explanations should be based on different effects for various categories.

5. Final Models

5.1 Model improvement

In order to solve previous questions, I improved the models from six aspects. Newly added variables will be introduced below. The descriptive statistics on the new variables are in table 5.

(1) Change to the ordinal logistic regression model

Happiness is measured by three ordered categories, ranked from “not too happy” (category 1), to “pretty happy” (category 2), to “very happy” (category 3). The structure of *happy* variable makes it more sensible to adopt ordinal logistic regressions rather than linear regressions for the following reasons.

While OLS assumes equal distance between categories, ordinal logistic regression sets an arbitrary spacing between categories. Ordinal logistic regression also explores the non-linear relationship between explanatory and target variables, opening more flexibility in modeling approaches. In addition, ordinal logistic models are not limited by the floor or ceiling effects because they treat the dependent variable as a cumulative probability. Lastly, the coefficients are more interpretable in ordinal logistic regressions than linear regressions. The improved model can inform the change in odds ratios of moving to the next category, in response to changes in independent variables.

When performing ordinal logistic regressions, one should note the parallel slopes assumption, which specifies a proportional odds ratio across category breaks. In other words, the model assumes that the odds of being in “not too happy” or “pretty happy” vs. “very happy” are the same ratio of being in “not too happy” vs. “pretty happy” or “very happy”. It also implies the same shape (and slope) of each logistic curve embedded within the larger model. The parallel slope assumption will be tested in a later section.

While *class* is treated as a continuous variable in the initial models, in fact *class* contains four exclusive categories with a rising order. The improved model recognizes the different class identifications and displays “lower class”, “working class”, “middle class”, and “upper class” separately. The categorical transformation of *class* presents different magnitudes of impacts on happiness brought by different social classes.

(2) Replace educ with a categorical degree

Initial models utilize *educ* to capture one’s educational background but yield minimal and insignificant results. To delve deeper into the education effect, I decide to change from “the number of schooling years completed” to “the highest degree obtained”.

While an additional year of schooling may have little impact on lifetime happiness, I expected a higher degree to result in a larger and more significant impact. Compared to the time duration of education, the highest degree reflects the quality of education, and thus distinguishes individuals on a higher level.

For instance, all else being equal, a Bachelor’s degree could qualify individuals for more prestigious and well-paid occupations than a high school degree. They could earn relatively high incomes and develop affluent and happy lifestyles.

degree is measured by five categories, including “Less than high school”, “High school”, “Associate/ Junior college”, “Bachelor’s”, and “Graduate.” As table 5 indicates, over half of the respondents have a high school degree. The next popular category is Bachelor’s degree, held by 20.92% samples. Proportions of “less than high school”, “junior college”, and “graduate” degree have similar magnitudes.

(3) Replace employed with a categorical wrkstat

employed in initial models is a binary variable that divides respondents based on whether they are currently employed or not. *employed* = 1 includes both those who are working full time and part-time. All other working status is classified as *employed* = 0.

In initial models, the variable *employed* generates different coefficients with varying significance. The negative effects of being *employed* also contradict my expectation. All such discrepancies signify the underlying complications within the employment indicator.

Therefore, to further examine the effects of employment on happiness, I further break down the working status into eight minor categories. As table 5 shows, "Working full time", "working part-time", "with a job but not at work because of temporary illness, vacation, or strike", "unemployed, laid off, looking for work", "retired", "in school", "keeping house", and "others." "No answer" and missing values are excluded from the model. Dissecting the binary categories allows one to capture the various responses to happiness for those in different working statuses. I expect to find significant differences in happiness by working status.

(4) Add prestige score to capture controls on occupation

I found the socioeconomic control variables in the initial models to be insufficient. To enhance the comprehensive list of controls, I added the prestige score of occupation. Specifically, I adopted the *prestg105plus* variable from the GSS dataset. It is an improved version of the original prestige score because it "removes the effect of individual raters." (GSS Codebook Appendix G) With a range from 6 to 97, the higher score indicates a more prestigious occupation. According to table 5, the prestigious score is approximately symmetric (skewness = $0.16 \in (-0.5, 0.5)$. Compared with normal distribution, *prestg* is less peaked distributed with shorter and lighter tails (kurtosis = $-1.26 < 0$). I expect to see a positive significant relationship between *prestg* and happiness with a 95% confidence interval.

(5) Considering interaction effects

As an intermediate process, I attempted some interaction terms to capture any underlying relationship among independent variables. For instance, I interacted income with the degree, prestige score, and working status respectively. I also attempted a similar series of interactions using *class*. I tried to add the square of *age* into the model. However, none of these interactions has generated significant effects on the dependent variable. Adding additional interactive factors does not improve the models' adjust R-square either. Instead, it renders the model even more difficult to interpret. Therefore, I decided to not pursue interaction terms in the final model.

Table 5. Descriptive Statistics on New Variables

Descriptive Statistics - New Continuous Variable

Statistics	N	Mean	St. Dev	Median	Min	Max	Range	Skew	Kurtosis
prestg	1377	49.83	26.12	50	6	97	91	0.16	-1.26

New Categorical Variable, Count and Proportion

Variables	Categories	Count	Proportion
degree_cat	Less Than High School	126	9.15%
	High School	693	50.33%
	Junior College	115	8.35%
	Bachelor's	288	20.92%
	Graduate	155	11.26%
class_cat	lower class	103	7.48%
	working class	630	45.75%
	middle class	598	43.43%
	upper class	46	3.34%
wrkstat	working full time	704	51.13%
	working part time	155	11.26%
	temporarily not working	27	1.96%
	unemployed, laid off	52	3.78%
	retired	271	19.68%
	school	39	2.83%
	keeping house	104	7.55%
	other	25	1.82%

5.2 Model Introduction: Ordinal Logistic Regression

Two final models are presented in table 6. The key independent variable in the first model is *ln_income*, and in the second model is categorical *class*.

Final Model 1: Association between Happiness Level and Income

The improved model 1 contains more significant variables than the initial model 1. Variables with statistically significant coefficients at 99% confidence interval are: *ln_income*, *prestg*, degree (junior college), and *good_health*.

ln_income: For one unit increase in people's *ln_income*, their odds ratio of being in one category higher (happier) is 1.401 times larger, controlling for all other explanatory variables.

prestg: For one point higher score of *prestg*, the odds ratio of being in one category higher (happier) goes up by 0.7% (= 1.007 - 1), holding all else constant. *prestg* has a significant relationship with *happy* on a 99% confidence interval.

degree_cat: the reference group is "less than high school degree." All categories of degrees display statistically significant differences in happiness probability from those who do not hold high school degrees, holding all else constant. Except for the reference group, the coefficients of all higher degrees are negative, indicating lower odds of being one category happier. By comparison, individuals with less than a high school degree have the highest odds ratio of being in one category happier, ceteris paribus. Specifically, for those with junior college degrees, their odds of being one category happier are lowest, with 0.482 times of those lower than high school degree. Among all degrees, the junior high degree has the most statistically significant relationship with one's overall happiness, with 99% confidence level. The negative relationship between degree and happiness makes sense to me. Consider an individual with a very low education degree as an example. If he/she can earn the same amount of income as those with much higher degrees, he/she is more likely to be happier.

Wrkstat: The model uses those working full time as the reference group. Among eight working status categories, some groups demonstrate significant relationships with happiness but in different directions. Individuals who work part-time and are unemployed in long term have negative coefficients. In other words, compared to the full-time working group, the odds ratio of moving one category higher (happier) goes down by 28.6% (0.714-1) and 43.6% (0.564 -1) for part-time workers and long-term unemployed people respectively, ceteris paribus. By comparison, the coefficients of retired and house-keeping groups are positive. This means that for the retired and house-keeping groups, the odds of being in one category higher (happier) are 1.488 and 1.703 times higher than full-time workers, ceteris paribus. The rest of the working status group (at school, temporarily not working, and others) do not exhibit significantly different patterns from full-time workers.

The diverging effects of working status on happiness can be explained by logic.

For the retired and house-keeping individuals, they do not bear the stress or working burden in the office but still have the same level of family income as the full-time workers. Thus, they may have more time to enjoy their lives and explore their personal interests. By comparison, part-time workers may be constrained by additional factors such as lacking advanced skills, which prevents them from engaging full-time work. They also tend to earn less than full-time workers, which may result in less happy states. In addition, long-term laid-off workers may feel socially excluded and thus more likely to be depressed than employed people. They may also suffer additional stress of finding suitable work.

age, female, white, good_health, relig: In the improved model 1, these variables do not exhibit different patterns from the initial model 1. The explanations on individual effects are as follows. For one year older reported, people's odds of being in a higher happiness level goes down by 0.7% (0.993-1), holding all other variables constant. Females' odds of being one category happier are 1.029 times higher than the males, holding all else constant. Whites' odds of being in one higher happiness level are 1.028 times higher than non-whites, ceteris paribus. Individuals's health effects are statistically significant at a 99% confidence level. Those with self-perceived good health have odds of being in a higher happiness category are 2.393 times larger than those suffering poor health, all else equal. Controlling for all other variables, having religious beliefs (vs. no religion) increases people's odds of one category higher (happier) by 29.5% (1.295 – 1). The religious effects are statistically significant at a 99% confidence interval.

Table 6. Summary of Final Models

Final Model: Ordinal Logistic Regression

	Dependent variable:	
	factor(happy)	
	(1)	(2)
ln_income	0.337***	
	(0.058)	
class_catworking class		0.625***
		(0.221)
class_catmiddle class		0.803***
		(0.228)
class_catupper class		1.509***
		(0.368)
prestg	0.007***	0.008***
	(0.002)	(0.002)
degree_cathigh school	-0.417**	-0.257
	(0.203)	(0.200)
degree_catjunior college	-0.730***	-0.514*
	(0.270)	(0.266)
degree_catbachelor	-0.474**	-0.263
	(0.242)	(0.237)
degree_catgraduate	-0.461*	-0.238
	(0.275)	(0.270)
wrkstatworking parttime	-0.336*	-0.407**
	(0.181)	(0.179)
wrkstattemp not working	0.245	0.241
	(0.388)	(0.388)
wrkstatunempl, laid off	-0.573**	-0.683**
	(0.291)	(0.293)
wrkstatretired	0.398**	0.232
	(0.191)	(0.189)
wrkstatschool	0.146	-0.037
	(0.337)	(0.333)
wrkstatkeeping house	0.532**	0.454**
	(0.213)	(0.215)
wrkstatother	0.084	-0.144
	(0.425)	(0.422)
age	-0.007	-0.007*
	(0.004)	(0.004)
female	0.028	0.012
	(0.110)	(0.110)
white	0.028	0.105
	(0.127)	(0.126)
good_health	0.873***	0.873***
	(0.129)	(0.130)
relig	0.259**	0.280**
	(0.127)	(0.127)
Observations	1,377	1,377
Log Likelihood	-1,269.106	-1,276.628
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 7. Odds Ratio vs. Coefficient Conversion Reference

Final Model: Ordinal Logistic Regression

	Coef	Odds_ratio	2.5 %	97.5 %
ln_income	0.337	1.401	0.224	0.451
prestg	0.007	1.007	0.002	0.012
degree_cathigh school	-0.417	0.659	-0.815	-0.020
degree_catjunior college	-0.730	0.482	-1.260	-0.203
degree_catbachelor	-0.474	0.622	-0.948	-0.001
degree_catgraduate	-0.461	0.631	-1.000	0.077
wrkstatworking parttime	-0.336	0.714	-0.691	0.017
wrkstattemp not working	0.245	1.278	-0.518	1.010
wrkstatunempl, laid off	-0.573	0.564	-1.146	-0.004
wrkstatretired	0.398	1.488	0.025	0.772
wrkstatschool	0.146	1.157	-0.516	0.809
wrkstatkeeping house	0.532	1.703	0.115	0.952
wrkstatother	0.084	1.088	-0.751	0.919
age	-0.007	0.993	-0.015	0.002
female	0.028	1.029	-0.188	0.245
white	0.028	1.028	-0.222	0.278
good_health	0.873	2.393	0.621	1.127
relig	0.259	1.295	0.010	0.508

Final Model 2: Association between Happiness Level and Class

Class_cat (categorical): Model 2 breaks the key independent variable *class* into four categories. The reference group is "lower class". All other three classes display significantly positive logit coefficients with a 99% confidence level, indicating higher odds ratios of being in one higher happiness category, ceteris paribus. The probability of feeling happier increases as one moves up the social ladder.

Specifically, working-class' odds of being in one category higher (happier) is 1.869 times higher than the lower class, controlling for all other explanatory variables. Holding all else constant, the middle class's odds ratio of one category happier is 2.232 times higher than the lower class. The upper class' have odds of being in one higher happiness category is 4.522 times of the lower class.

Besides, there are a few other significant variables in predicting happiness at a 99% confidence level, including *prestg* and *good_health*. working part time, long term unemployed, keeping house, and *relig* are significant on 95% confidence level. degree of junior college and age have coefficients with 90% statistical significance.

Compared with final model 1 on income effects, model 2 displays similar socioeconomic effects regarding magnitudes and directions. Thus, the interpretation of coefficients can be similar as discussed above. However, I hope to point out several differences across the two final models. First, degree becomes a less significant predictor of happiness in the *class* model. This could be partially due to the high correlation between class and degree level. The degree effects on happiness are largely explained by class, leaving little information reflected in degree categories. By contrast, family income may explain less of the degree effects on happiness. Because family income is a composition of every member's wages, it does not directly link to inidivual's earning or educational background. Therefore, even with the presence of family income in the model, degree can exhibit a significant relationship with one's general happiness.

In addition, *age* has the same coefficient in the income and class models. However, it becomes a significant predictor of general happiness with a 90% significance level. For each one year older in age, people's odds of being in one happier category goes down by 0.7% (0.993 – 1). The difference in significance level may be explained by the lower correlation between one's class and age. Specifically, people could be born as upper class or doomed to be a lower class for their entire lives. Social immobility fixes individual class status. By contrast, the family income could be more correlated with age. As people grow older, their family income accumulates. Therefore, compared to class, the family income could explain more of the age effects on happiness. As a result, when family income is present in the model, age becomes insignificant predictor of income.

Table 8. Odds Ratio vs. Coefficient Conversion Reference 2

Final Model: Ordinal Logistic Regression

	Coef	Odds_ratio	2.5 %	97.5 %
class_catworking class	0.625	1.869	0.193	1.059
class_catmiddle class	0.803	2.232	0.358	1.250
class_catupper class	1.509	4.522	0.793	2.240
prestg	0.008	1.008	0.003	0.013
degree_cathigh school	-0.257	0.773	-0.649	0.134
degree_catjunior college	-0.514	0.598	-1.037	0.008
degree_catbachelor	-0.263	0.769	-0.729	0.202
degree_catgraduate	-0.238	0.788	-0.769	0.291
wrkstatworking parttime	-0.407	0.665	-0.760	-0.057
wrkstattemp not working	0.241	1.272	-0.522	1.005
wrkstatunempl, laid off	-0.683	0.505	-1.260	-0.109
wrkstatretired	0.232	1.261	-0.139	0.603
wrkstatschool	-0.037	0.964	-0.693	0.617
wrkstatkeeping house	0.454	1.575	0.033	0.878
wrkstatother	-0.144	0.866	-0.973	0.686
age	-0.007	0.993	-0.015	0.001
female	0.012	1.012	-0.204	0.228
white	0.105	1.110	-0.143	0.353
good_health	0.625	1.869	0.620	1.128
relig	0.803	2.232	0.032	0.528

5.3 Result Diagnostic

In order to test the parallel slope assumption, I performed the Brant test on *ln_income* and *class* respectively. The null hypothesis states that slopes are identical across levels of key independent variables.

As the result shown in table 9, the parallel slope assumption is violated in both regressions. The P-values for both regressions are below 0.05, and thus we reject the null hypothesis of parallel slope. In short, the test finds evidence for a significantly different slope for each level of *ln_income* and *class*, with a 99% confidence level.

Table 9. Brant Test on Parallel Slopes

Brant Test on Parallel Slopes						
	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
Test on ln_ncome						
lm1_brant	19	2611.1	-1286.5			
lm1	20	2578.2	-1269.1	34.873	1	3.52e-09***
Test on class						
lm2_brant	19	2611.1	-1286.5			
lm2	22	2597.3	-1276.6	19.829	3	0.0001842***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

6. Conclusion

6.1 Conclusion

This study discovers both family income and social class have significant positive relationships with general happiness. Individuals with higher family incomes are happier than those with lower family incomes. In addition, perceiving oneself in a higher social class increases one's happiness level. Thus, both objective economic resources and subjective social status influence subjective well-being. The results match my hypothesis. In addition, this study also finds other factors to be significantly related to general happiness, including educational degree, working status, occupational prestige, religious belief, age, and health condition. These factors represent a comprehensive view of individual's socioeconomic status in impacting happiness.

6.2 Limitation

The initial linear regression models are insufficient due to untested assumptions and difficult interpretation. However, the improved ordinal logistic regressions also contain limitations.

First, the parallel slope assumption is violated. The test indicates different slope coefficients on income and class for each threshold jump. Future studies should consider multinominal logistic regression or simple binary logistic regression to address such violations.

Second, the measurement approach of family income is lacking. GSS survey records family income by brackets. While the midpoint of each bracket serves as an approximation of actual family income, it still lacks precision compared to the actual dollar value of aggregated income. An appropriate imputation method on extreme income groups also requires further investigation.

Third, this study only examines cross-sectional survey responses in 2018. Because one-year income data may contain temporary fluctuation, future studies should consider using an average of long-term family income measures.

6.3 Project Reflection & Extension

This project covers a broad range of models we learned this semester. I learned the importance of encoding variables into their suitable formats. For instance, the initial model encoded social class on a numerical scale, rendering a difficult interpretation of the result. By comparison, the final model disaggregates class into four categories, and encodes as dummy variables. Such a recoding technique improves the result significantly. In addition, I learned the importance of understanding the variables before building the model. For instance, there are several income-related variables in the GSS dataset. While they all capture the income information, these variables differ in their focus on different aspects. In order to select the most relevant variable, I need to consider the research design, modeling structure, and the purpose of this project.

While the current model comprises limitations, I would love to pursue this topic further and integrate more advanced modeling approaches in my master's thesis. In addition, as GSS also contains a panel dataset, longitudinal research is another potential extension of my current progress. As a next step, I aim to explore the short-term dynamic of happiness level. By comparing an individual's happiness over time, I plan to associate such changes with their life events.

Reference

Firebaugh G, & Tach L. (2012). *Social trends in American life: findings from the General Social Survey since 1972. Chapter 10: Income, Age, and Happiness in America*.

GSS Codebook Appendix G. (n.d.). Retrieved December 14, 2021, from https://gss.norc.org/documents/codebook/GSS_Codebook_AppendixG.pdf

Hout, M., & Greeley, A. (2012). *Social Trends in American Life: Findings From the General Social Survey Since 1972. Chapter 11: Religion and Happiness*. https://web-s-ebscohost-com.ezproxy.cul.columbia.edu/ehost/ebookviewer/ebook/ZTAyNXhuYV9fNDgyMTQ4X19BTg2?sid=f50016bb-9d05-4456-8fee-00c545ac8611@redis&vid=0&format=EB&rid=1

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(38), 16489. https://doi.org/10.1073/PNAS.1011492107

Ligon, E. (1989). *The Development and Use of a Consistent Income Measure for the General Social Survey*.

Mistry, R. S., Elenbaas, L., Griffin, K. M., Nenadal, L., & Yassine, A. (2021). Advancing developmental intergroup perspectives on social class. *Child Development Perspectives*, *15*(4), 213–219. https://doi.org/10.1111/CDEP.12431

Stevenson, B., & Wolfers, J. (n.d.). *The Paradox of Declining Female Happiness**. Retrieved December 16, 2021, from http://bpp.wharton.upenn.edu/betseyshttp://www.nber.org/~jwolfers