The Economics of Psychological Well-Being: Evidence From the United States Using Machine Learning Method

Yanchao Yang*

Abstract

This paper models peoples psychological well-being (PWB) by applying machine learning methods to a large dataset in the United States. PWB can be quantified by three self-reported variables: general happiness, satisfaction with the financial situation, and satisfaction with the job. First, we run the K-nearest neighbor (KNN) algorithm for each respective feature to rank the importance of features. We found that marital status has a higher importance score than income for peoples general happiness. Prestige score of occupation is the most important predictor of satisfaction with jobs. Next, we utilize the Forward Selection algorithm to find the best combination of predictions. Using this selected combination to predict people's PWB, we achieve 70% - 80% classification accuracy when detecting people with disadvantaged PWB. On top of that, we use ordered probit regression to quantify how each feature affects PWB. Lastly, we show that PWB is important to economics by investigating how happiness affects physical and mental health, risky goods consumption, investment decisions, and working behaviors. We find that happier people have better health conditions, smoke and drink less, have more confidence in financial institutions, and generally work more hours.

Keywords:Psychological Well-Being, Machine Learning

JEL Classification: A12, B55, C10, D91, E29

*Department of Economics, 3124 Sproul Hall, University of California, Riverside, CA 92507, US. Email:yyang056@ucr.edu

1 Introduction

Happiness, or psychological well-being(PWB), is considered an ultimate goal of life. The United States Declaration of Independence of 1776 takes it as a self-evident truth that the pursuit of happiness is an unalienable right, comparable to life and liberty(Frey and Stutzer, 2002). However, ever after World War II, the psychological well-being in the United State has been flat in line with considerable increases in economic growth and personal income(Easterlin, 1974). Even with the rapid growth of economics and PWB literature over the past 50 years, there still remain many open questions in this field.

This paper has two aims. The first is to find the key features to predict people's psychological well-being and then predict and classify the different state of people's PWB. The second is to study how the extent of PWB may influence many economic decisions and behaviors. In detail, this paper uses the large individual-level dataset drawn from the General Social Survey(GSS) that covers a long time span(1972-2018) and broad socioeconomic topics in the United States. The key measurements of psychological well-being are people's general happiness, satisfaction with financial situation, and satisfaction with jobs. This paper adopts machine learning approaches, K-Nearest Neighbor(KNN) algorithm and Forward Selection algorithm, to detect the most important features that affect PWB. Then use the selected features to predict individual well-being. We use the ordered probit model to see how each feature relates to PWB. In addition, we study the potential impacts of PWB on physical and mental health, consumption activities, working behaviors, and investment behaviors.

This paper is closely related to two strands of the economics of PWB study. The first focuses on how economic policies and the institutional condition affects peoples PWB. Most notably, Easterlin(1974) found a small linkage between happiness and GDP per capita. A more recent study(Helliwell et al., 2012) found similar results for many other countries. Unemployment is another individual attributes that negatively affect PWB (Clark and Oswald, 1994; Winkelmann and Winkelmann, 1998; Grn et al., 2010). Many studies found the racial difference in PWB gap in the U.S.(Deaton and Stone, 2016) and the gap is also shrinking over time(Stevenson and Wolfers, 2009).

The second strand of literature considering PWB as the explanatory variable and see how happiness affects peoples consumption, investment, and human capital, etc,. Oswald et al. (2015) uses experiments to show that happiness increases productivity. Happier people are also more likely to be healthier, (Danner et al., 2001), have better immune systems, less inflammation, and fewer infections(Epel,2009). Unhappiness, on the other hand, is related to risky behaviors such as smoking(Brandon, 1994), drinking, and marijuana use(Magid et al., 2009). Lucas et al. (2003) also found people with higher levels of life satisfaction are less likely to divorce or separate. Cetre et al. (2016) showed that happiness is important in predicting future marriage and fertility.

There are three main shortcomings of the current literature. First, the cross-person comparisons of subjective feelings such as "happy" or "satisfied" are likely to be unreliable because there is no natural scaling to do the comparisons. Second, most of the literature focuses on the impact of one particular factor on PWB, but happiness is jointly determined by many different aspects of life. When people are facing a bundle of choices, they need to know which factors attributes most to their happiness. Yet little study focuses on this question so far. Third, there is no study that provides a reliable method to detect people with disadvantaged psychological well-being. People who are unhappy are more likely to have depression, anxiety, and other mental illness. Early intervention could be more effective. Policy makers should focus on investigating people's psychological health and implement early assistance.

This paper contributes to the literature in the following three ways. First, this is the first paper focusing on the classification of PWB using machine learning methods, which help us to detect individuals who are at a higher risk of depression and guide interventions to assist them. The selected features are common variables in many databases, making it possible for future researches to conduct the out-of-sample prediction of PWB. Second, this is the first paper that finds the importance of prestige of occupation on PWB. This finding adds strong evidence to psychological and economic literature that self-esteem and social-value are very crucial. Third, this is the first paper that studies the two stages of psychological well-being. The classification of the state of psychological health and the impacts of PWB on economic activities.

The main results of this paper are as follows. The first finding is that marital status is the most important feature that relates to individuals' general happiness. The prestige of occupation, which has long been ignored from the current literature, come out to be significantly crucial to satisfaction with financial situations and jobs. Income is also a strong predictor for PWB. In addition, the KNN algorithm performs well in detecting people with disadvantaged psychological well-being status. We then use the ordered probit regression model to see how each of those features affect PWB. We found that happier people tend to be those who have no child, with higher income, more prestigious jobs, and married individuals. The second stage of this paper finds that unhappy individuals are more likely to consider suicide if negative events happen in their lives. They are more likely to have an HIV test and generally have 3 more days a year with poor mental health. Unhappy people also smoke and excessively drink more. One the other hand, people who self-reported as happy are having better health, more confident in the financial institutions, and banking system, which implies a more active potential investment behavior.

The rest of the paper is organized as follows: Section 2 summarizes the related literature. Section 3 describes the details of the dataset. Section 4 introduces the empirical methodologies we implemented. In section 5, we present the empirical results. Section 6 shows how PWB relates to economic activities. Section 7 concludes.

2 Literature Review

The study of psychological well-being(PWB) and economics begin in the 1970s and has been drastically grown over the past 50 years. There are two mainstreams of the study. As Frey and Stutzer(2002) reviewed, the first strand focus on how economic policies, and the institutional condition affects peoples PWB as well as the formation of PWB. The second body of literature considering the PWB as the explanatory variable and to see how happiness affects peoples consumption, investment, and human capital, etc,(Magid, Colder, 2009; Edmans, 2012; Chen et al., 2020; Labroo et al., 2009).

Economists have found evidence that PWB is systematically related to both the individual characteristics and economic and social aggregate characteristics. One of the most prominent studies is done by Easterlin(1974). He found a small linkage between happiness and GDP per capita, with the personal income grew over time while peoples self-reported happiness doesnt increase. A more recent study(Layard and Sachs 2012) used the Gallup World Poll data and found similar results for many countries. To explain this puzzle, Kapteyn et al. (1997) focused on how the preference changes due to social comparison. Bartel(1981) studied how relative income affects PWB by checking the racial difference in satisfaction with job. Using either cross-sectional data or panel data, unemployment is another individual attributes that negatively affect PWB (Clark and Oswald,1994; Winkelmann and Winkelmann,1988; Grun et al., 2010). Age is found to be U-shaped related with PWB(Blanchflower and Oswald, 2007). Gender and race are also important indicators for PWB. Many related studies failed to disentangle the confounding impact of the labor market outcomes that substantially exists across different gender and race. Overall, women have higher satisfaction scores for their life and jobs(Clark 1997). Though the gender PWB gap is noted to shrink in many countries(Stevenson and Wolfers,2009). Many studies found the race difference in PWB gap in the U.S.(Deaton and Stone,2016) and the gap is also shrinking over time(Stevenson and Wolfers,2009). Marital status is also related to PWB. Married people usually report higher happiness scores. Divorce has prolonged negative impact on people's happiness. Frey and Stutzer(2006) found the reverse relationship, where it is happier individuals that are more likely to get married. The finding on how education impact PWB is mixed. Di Tella et al. (2001) found that education is monotonically related to happiness scores. But it is difficult to find the net effect of education on happiness because education raises peoples income and other expectations.

Macroeconomics conditions are systematically related to peoples happiness too. People obtain information about the macroeconomic variables regularly from newspapers or social media, which suggests that aggregate economic conditions matters to peoples feeling. Tella et al,(2001) used European data found country-level correlations between happiness and GDP per capita, aggregate unemployment, and inflation. They found large psychic loss due to the recessions. In a later article, they estimated the trade-off between unemployment and inflation using misery index and they found that unemployment have larger negative effects on PWB than inflation. Blanchflower et al,(2014) revisited this topic by using the updated European data and found similar results that the impact of unemployment is 5 times larger than aggregated inflation in lower the well-being.

The literature on the impact of psychological well-being on peoples behaviors and choices has also growing rapidly. Oswald et al., (2015) using experiments by giving people some happy stimulus and they showed that happiness increases productivity. De Neve et al., (2013) found that PWB is a strong predictor of future earnings. Happier people are also more likely to be healthier, (Danner et al., 2001), have better immune systems, less inflammation and fewer infections(Epel,2009). Unhappiness, on the other hand, is related to risky behaviors such as smoking(Brandon,1994), drinking and marijuana use(Magid and Colder,2009). Clark et al., also found people with higher levels of life satisfaction are less likely to divorce or seperate. Cetre et al.(2016) showed that happiness is important in predicting future marriage and fertility.

Individuals psychological well-being is being jointly determined by many different factors. Personality traits, income, education, and the macroeconomic conditions. To the best of our knowledge, our paper is the first one that use the machine learning methods to analyze the determinants of happiness in different phrases of business cycle. The categorical nature of the PWB data enables us to leverage the machine learning algorithms more efficiently.

3 Data

3.1 Psychological Well-Being

In this paper, I use the data drawn from General Social Survey(GSS), which was conducted by the National Opinion Research Center. GSS data is available between 1972-2018. The survey was conducted almost every year between 1972-1991 and then every other year between 1993-2018. Every year they GSS interviews around 2000 individuals, which brings up the total number of observations more than 60,000 over all the timespan. Questions they asked in the questionaires are very broad, including the information about the respondents' demographics, financial conditions in their households, their point of view about social, cultural, and political issues. I use the self-reported data as the measurement of psychological well-being. The variables used are the following: respondents' general happiness, satisfaction with financial situation, and job or housework.I focus on these variables because they were available in the 31 survey waves from 1972-2018. The questions being asked are the follows:

> **General happiness:** "Taken all together, how would you say things these days? Would you say that you are very happy, pretty happy, or not too happy? 1) Very happy; 2) Pretty happy; 3) Not too happy."
>
> **Satisfaction with financial situation:** "We are interested in how people are getting along financially these days. So far as you and your family are concerned, would you say that you are pretty well satisfied with your present financial situation, more or less satisfied, or not satisfied at all? 1) Satisfied; 2) More or less; 3)No at all satisfied."
>
> **Satisfaction with Job:**"On the whole, how satisfied are you with the work you do–would you say you are very satisfied, moderately satisfied, a little

dissatisfied, or very dissatisfied? 1) Very satisfied; 2) Moderately satisfied; 3) A little dissatisfied; 4) Very dissatisfied."

The raw happiness descriptive statistics are presented in Table 1. Overall, respondents' say they are fairly happy or very happy and the distribution skewed towards the top of the distribution. There is no much difference across gender. People who experienced unemployment and divorce do report higher percentages of unhappy. Income is also crucial. The percentages of not happy answers decreases as the income quartiles increases.

Both of the household and personal income are provided by 12 categories in the raw data. In order to convert the categorical income data into continuous variable, I obtained the data from Current Population Survey (CPS) and calculate the mean and standard error for each categories by year. Then randomly assign the value to each respondents by survey year according to normal distribution. I adjust all income variables in 2012 dollars.

Table 1: Happiness in the United State:1972-2018

			Marital Status	
Self-Reported Happiness	All (%)	Unemployed (%)	Married (%)	Divorced (%)
Very Happy	31.34	21.24	40.44	19.62
Pretty Happy	55.89	54.12	51.91	61.38
Not Too Happy	12.77	24.64	7.65	19.00

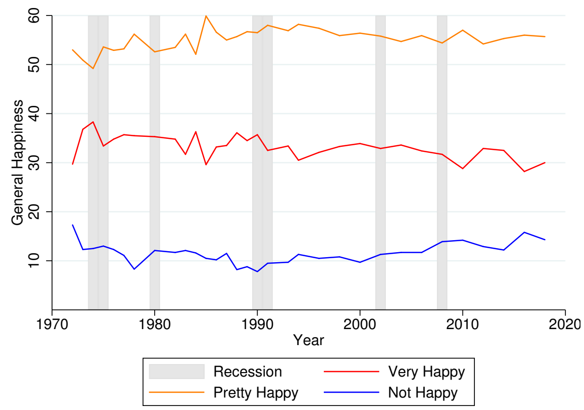
	Sex		Income Quartiles			
Self-Reported Happiness	Male (%)	Female (%)	1st (Lowest)	2nd	3rd	4th (Highest)
Very Happy	30.70	31.85	22.53	24.23	26.96	33.62
Pretty Happy	56.64	55.29	52.09	55.69	56.34	56.74
Not Too Happy	12.66	12.86	25.38	20.08	16.71	9.64

Note: the above descriptive statistics are based on 60054 observations.

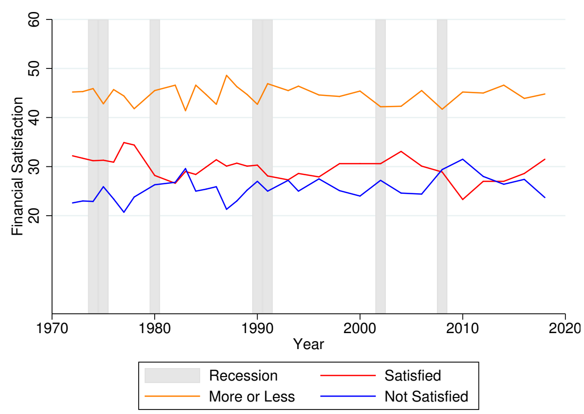
3.2 Macroeconomic Data

Tella, MacCulloch and Oswald(2003) have shown that macroeconomic movements have strong effects on people's happiness. To take the macro-level shocks into consideration, I also merge the following macroeconomic variables: annual real gross domestic product per capita, real personal expenditures, unemployment rate, inflation rate, and recession indicators. These variables are drawn from the Federal Reserve Economic Data (FRED) website of the Federal Reserve Bank of St. Louis. All money related variables are chained into 2012 dollars.

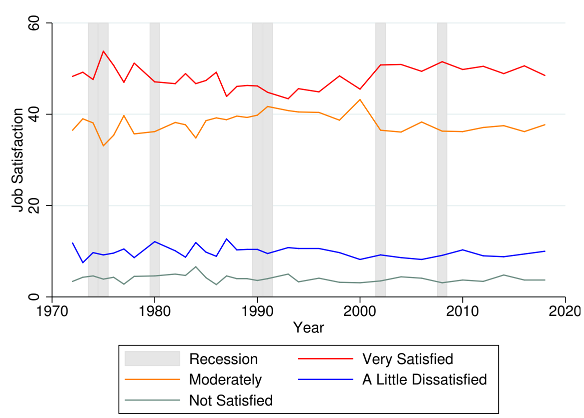
We graph the PWB trends in Figure 1. The overall PWB trends are stable, even though the real GDP per capita have grown in the U.S. for those decades. As similar results that Easterlin(1974) found no evidence that the happiness data are trended over time. The shaded areas are the recession time. The trends shows some patterns during the business cycles. With the percent of ”not happy” and ”not satisfied” respondents increases during the great recessions in 2008.



(a) General Happiness



(b) Satisfaction with Financial Situation



(c) Satisfaction with Job

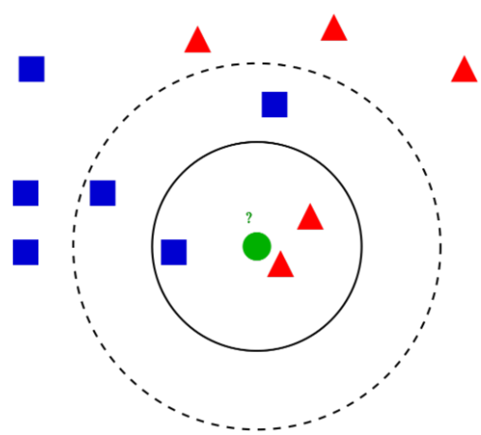
Figure 1: Trends in Psychological Well-being

4 Empirical Methods

4.1 K-Nearest Neighbor(KNN) Algorithm

K-NN is one of the most fundamental non-parametric algorithms and it has been widely used for classification in public health and clinical studies. K-NN is also one of the supervised machine learning algorithms. Generally, K-NN is used in two different ways: K-NN classification and K-NN regression. In this study we are going to apply K-NN to classify people's different state of PWB. To determine or classify an instance, K-NN will see what is the majority of the K-nearest neighbors of this instance. For example (in Figure 2), there are two red instances and one blue instance out of three nearest neighbors of the green instance. Thus the green instance will be marked as red according to our K-NN prediction. If the prediction was right, it will count as a successful prediction, otherwise a failed prediction.

Figure 2: KNN Illustration



From the raw dataset, a 16-parameter vector describes each respondent. "General Happiness", "Satisfaction with the financial situation", "Satisfaction with the job" are three picked parameters as our outputs of the K-NN algorithm, and the rest 13 parameters are used as input data. Each output will be matched with the rest 13 parameters to form a 14-parameter vector, which means we are running three subsets in our experiments. This algorithm works based on the distance between a test sample and specified training samples (Peterson,2009). The distance metric is important when implementing KNN. In this paper, we use the Euclidean distance function. Let x_{ij} to represent the input features with n $(i = 1, 2, \ldots, n)$ number of observations and $f(j = 1, 2, \ldots, f)$ number of features. The Euclidean distance between input features x_i and output class x_c is defined as:

$$d(x_i, x_c) = \left(\sum_{i,c=1}^{n} (|x_i - x_c|)^2 \right)^{\frac{1}{2}} \tag{1}$$

Since this algorithm relies on distance for classification, we need to normalize our data before implementing a classification algorithm. We normalize the training data by re-scaling predictors to [0,1] to improve the classification accuracy. In addition, we assigned weights to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. This is called “The weighted nearest neighbor classifier ”. In this study, we add weights by assigning each neighbor a weight of 1/r, where r is the Euclidean distance to the neighbor.

One of the main drawbacks of K-NN is its sensitivity of outliers or irrelevant features. If we don’t remove those outliers and irrelevant features, the classification accuracy can be dramatically degraded. Therefore, it is a kind of algorithm thats very sensitive to irrelevant variables. The accuracy of the classification drops in a great deal if there are irrelevant features in the model. Then how do we mitigate the nearest neighbor algorithms sensitivity to irrelevant features? Typically there are three ideas: 1. Use more training instances; 2. Use statistical tests to try to determine which features are useful; 3. Search over feature subsets. In this study, we applied a search algorithm which is Forward Selection Search to solve this problem and try to find out the combination of features that give us the best classification rate.

In forward selection, the whole procedure is doing a forward single variable selection which approaches a higher success rate. The first variable selected for an entry into the constructed model is the one with the largest correlation with the dependent variable. Once the variable has been selected, it is evaluated on the basis of certain criteria. The criteria here is to see if the prediction accuracy is the highest. If the first selected variable meets the criterion for inclusion, then the forward selection continues. The procedure stops, when no other variables are left that meet the entry criterion (Walczak and Massart, 2000). During the procedure, if a higher success rate can be obtained, the same process is repeated once again retaining the two selected features and adding a third one, one at a time, until all remaining features have been used. The process is then iteratively repeated until no better combination can be obtained. The result of this procedure is a series of features that represent the best multivariate combination.

Based on these traits, we can use KNN to rank the importance of different features and find how multiple features combined would determine happiness. Unlike probit regression models or linear regression models that many researchers have used to analyze people's happiness(Tella, MacCulloch and Oswald,2003; Jackson,2017), KNN helps us to find the combined features that contribute most to one's psychological well-being. So we can provide a systematic analysis on the determinants of PWB.

4.2 Model Performance Evaluation

When the sample is biased towards a certain features in some way, the overall accuracy rate might also be biased. This is why we use the confusion metrics other than overall accuracy to evaluating a machine learning algorithm. By tabulating each of the of predicted and true value, we can evaluate the accuracy rate for each class.
A general way of constructing a confusion matrix is the following: we use TP to

denote the true positives, which means when the outcome is positively and is predicted as positive. When a negative outcome is predicted as positive, we denote this case as false positives(FP). Similarly, we define true negatives(TN) as the predicted and actual outcome are all negative. When the actual results are positive but we predict it as negatives, we call it false negatives(FN). Then, we use the following measurements to evaluate the model performances:

$$TPR = \frac{TP}{TP + FN}$$

$$TNR = \frac{FP}{FP + TN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

The true positive rate(TPR) measures the proportion of correctly predict the outcome when it is positives which is also referred to as sensitivity or recall. The true negative(TNR) measures the proportion of negative outcomes that are called negatives. In this study, we pay more attention to the true negative rate

$$F_1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

We also use $F_1 - score$ to calculate the balanced accuracy which is calculated by the weighted harmonic average of precision and recall. We present these results in the confusion matrices.

4.3 Ordered Probit Regression

KNN helps us to find the most relevant features. To see how each of those features affect people's psychological well-beings, we use the ordered probit model to deal with the categorical data using the following equation, where PWB_{ijt} is the measure of psychological well-being j by individual i in year t. $PersonalTraits$ includes a set of variables that include the income, sex, marital status, education, number of children, working status and age. λ_t is the year fixed effects and u_{ijt} is the unobservables.

$$PWB_{jit} = \beta_j KeyIndependentVariables_{jit} + \sum \alpha_j PersonalTraits_{jit} + \lambda_t + u_{jit} \quad (2)$$

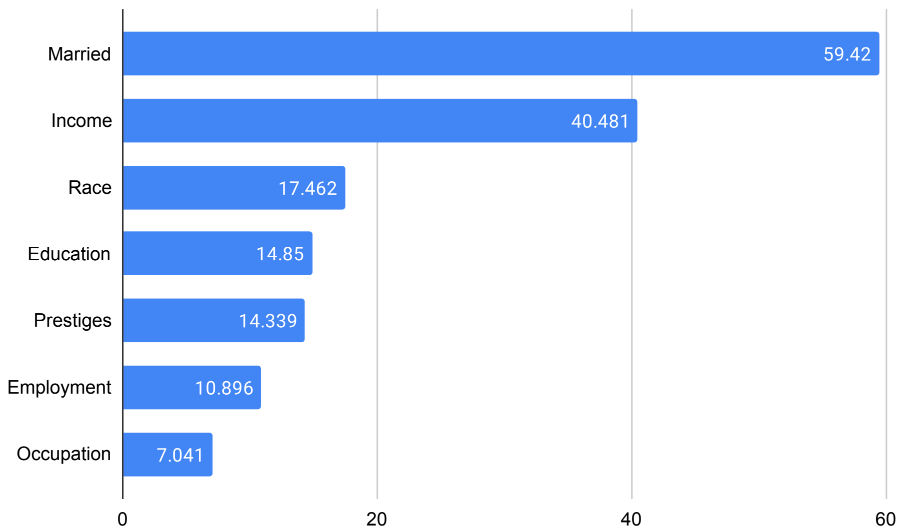
The results are shown in Table 5 and the results including the year fixed effects. Consistent with what we have found using KNN methods, two features stand out: unemployment and marriage. Joblessness have the largest and negative effect on individuals' satisfaction with financial situation. Marriage is positively related with all three measurements of PWB and has largest impact on the likelihood that a respondent says he or she is happy about life. People's PWB also increasing with the higher income, but increasing with a decreasing rate as the impacts gets smaller from the third quartile to fourth quartile. The effect of is monotonic. We also detect a life cycle patterns in PWB, which is U-shape in age.

5 Empirical Results

5.1 Feature Importance Ranking Using KNN

In this section, we use KNN to determine the variables used in classification process by calculating the importance scores of each variable. Figure 3 shows that marital status is most important in predicting people's general happiness. Income ranked second. We can notice that the prestiges scores of people's occupation is also crucial in predicting happiness. The prestiges scores can bu understood as people's occupational reputation. For example, physicians, professors in universities, and lawyer have the highest prestiges score. As some literature documented, people's happiness is related to self-esteem and the social comparison. So among al job–related variables(employment status and type of occupation), prestiges exhibits the higher importance.

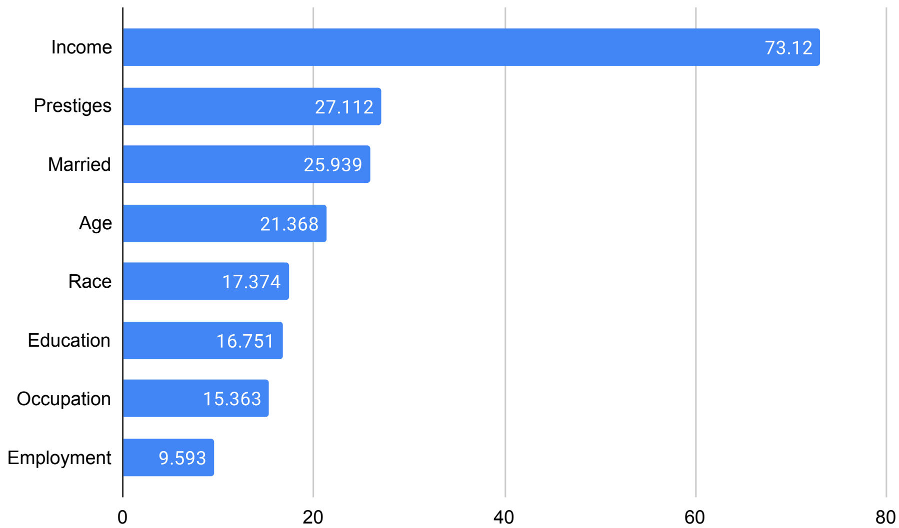
Figure 3: The Importance Score for General Happiness



Note: the importance scores are scaled to 100.

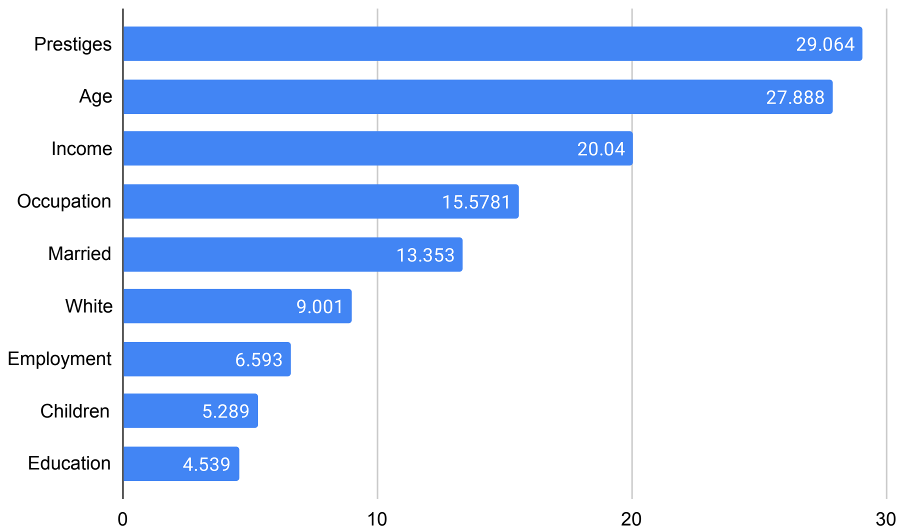
Figure 4 shows the importance of ranking of respondents' satisfaction for financial situation. Real income is most important. Next is prestiges score. Income and prestiges score are also closely related. People with higher prestigious jobs tend to make more money. Marital and age are also crucial predictors. Figure 5 presents the importance score for satisfaction for job. Prestiges and age are most important features. The number of children turns out to be important. The balance between work and parenting is always an important topic in social science. Using the probit model, we found that having no child is negatively related to job satisfaction.

Figure 4: The Importance Score for Satisfaction for Financial Situation



Note: the importance scores are scaled to 100.

Figure 5: The Importance Score for Satisfaction for Job



Note: the importance scores are scaled to 100.

5.2 Classification Findings

In this section, we investigate the the performance of KNN algorithm in classifying people's state of general happiness, satisfaction with financial situation and satisfaction with jobs using the selected features. We present the results based on the confusion matrices that clearly show the precision rate and recall rate.

Firstly, we show the results for general happiness. When implementing KNN algorithm, picking the appropriate value of K is very crucial in order to avoid overtraining and oversmoothing.We divide the dataset into training set and testing set to detect these two problems. By experimenting different values of K, we found that the accuracy rate converges at $K = 7$. So the following results are shown for this particular K value. By calculating the feature importance using KNN, we manage to pick the variables that are most relevant to general happiness. The features we use in the classification are the following: the marital status, the real household income, race, the highest years of school, the employment status and the prestiges score of their occupation.

As shown in Table 2, the KNN algorithm performs best in classifying unhappy respondents, with the rate of 75.9%. Our model performs weakly in predicting “pretty happy” people with the correction rate of only 38.9%. The rate of correctly predicting“very happy” people is 60.9%. People only choose “not happy” when they are really not happy, which is the common issue for self-reported psychological data, especially with 3 possible answers. Most respondents pick the moderate(middle) answer if they don’t have strong opinions. Therefore, the prediction using the available features tends to be vague for the “pretty happy” individuals.

From the perspective of public policy, being able to detect “not happy” individuals is more important. As shown in the existing study, unhappy people are more likely to smoke more(Coan,1973; McKennel,1970; Shiff- man,1993; Becona, Vazquez, Lorenzo,1998)), drink more alcohol(Magid, Colder, and et al.,2009), more likely to have mental physical problem(Watson and Pennebaker,1989; Lagdish,1993;Curhan, Sims, and etc.,2014). Our model provides policy-makers with an feasible implementation to target unhappy people and make corresponding policies to promote the overall social welfare.

Next, we present the results for people’s satisfaction with financial situation. The features we use are working status, age, marital status, race, educational level, real household income, religion type, prestiges score of their occupation. The value of K is 5. People’s satisfaction with their personal financial situation is an important indicator in predicting their future investment and consumption behaviors. People with higher satisfaction level are generally more likely to consume more and have more investment diversity. Our results shows that the KNN algorithm performs best for individuals who are not satisfied with their financial situation at all with recall rate 73.72%. Our model performs weakly for people whose opinion are neutral. The prediction of “pretty well” is 65.51%.

Table 2: Classification for General Happiness

True	Classified Not Happy	Pretty Happy	Very Happy	Total
Not Happy	4566 75.9%	613 10.19%	837 13.91%	6016 12.34%
Pretty Happy	8033 29.19%	10704 38.90%	8780 31.91%	27517 56.48%
Very Happy	3285 21.63%	2654 17.47%	9251 60.90%	15190 31.18%
Total	15884 32.6%	13971 28.67%	18686 38.73%	48723 100%
Priors	0.3333	0.3333	0.3333	

Note: The first rows of each cell are the number of observations being classified into the corresponding categories. Results are based on weighted KNN

The results for classification for job satisfaction is shown in Table 4. The predictors including respondents' working status, age, number of children, marital status, race, education, real household income and occupation. The job satisfaction could be an indicator that predicts people's job quit probabilities. Cote and Morgan(2002) found that the decreases job satisfaction would increases the intentions to quit. The model performs relatively well for individuals who are not satisfied with their job at all.

Overall, the KNN algorithm performs well in classifying people who are unhappy or dissatisfied with their jobs and financial situation. People who are in a less advantaged state of psychological well-being might have more common characteristics, which makes the predictors to be stronger in predicting the outcomes. In the next section, we use probit model to show how each feature affects people's psychological well-beings.

Table 3: Classification for Satisfaction with Financial Situation

True	Classified Not at all	More or less	Pretty well	Total
Not at all	9694 73.72%	1688 12.68%	1788 13.6%	13150 27.02%
More or less	6537 30.06%	8966 41.24%	6240 28.7%	21743 44.67%
Pretty well	2953 21.31%	1827 13.18%	9080 65.51%	13860 28.48%
Total	19184 39.35%	12461 25.56%	17108 35.09%	48673 100.00%
Priors	0.3333	0.3333	0.3333	

Note: The first rows of each cell are the number of observations being classified into the corresponding categories. Results are based on weighted KNN

Table 4: Classification for Satisfaction with Job

True	Classified Not at all	Moderate	Very satisfied	Total
Not at all	3510 62.9%	1180 21.15%	890 15.95%	5580 14.21%
Moderate	3190 21.1%	7640 50.53%	4290 28.37%	15120 38.52%
Very satisfied	4090 21.27	4950 25.74%	10190 52.99%	19230 48.98
Total	10790 27.02%	13770 34.49	15370 38.49	39257 100%
Priors	0.3333	0.3333	0.3333	

Note: The first rows of each cell are the number of observations being classified into the corresponding categories. Results are based on weighted KNN

5.3 Results of Probit Regression

We use probit regression to see how each features affect people's psychological well-beings. We found some interesting and similar patterns for these three subjective well-being variables. Panel A of table 5 show that people tends to be happier and more satisfied with their financial situation. Having three or more kids would negatively and significantly affect financial satisfaction. In panel B, income is monotonically related to all three outcome variables. Panel C shows that unemployment have large and statistically significant impact on people's well-being. Self-employed people seem happier. Marriage has large impact on people's happiness. The causality link between happy and marriage is bidirectional as documented in literature(Frey and Stutzer,2006;). Getting married makes people happier, more satisfied with their financial situation or jobs. It is likely that happier people are more likely to get married. In all, happier people seems to be those who have no kid, higher income, more prestigious job and married ones.

Table 3: Psychological Well-being and Personal Characteristics: Full Sample Ordered Probit Model

Independent Variable	General Happiness	Satisfaction with financial situation	Satisfaction with Job
Panel A: Number of Children	(1)	(2)	(3)
No child	0.0595***	0.155***	-0.0743***
	(0.0146)	(0.0144)	(0.0168)
One child	-0.0437**	-0.0468***	0.00128
	(0.0143)	(0.0141)	(0.0162)
Two children	0.00305	0.01	0.0182
	(0.0124)	(0.0122)	(0.0143)
Three or more	-0.0119	-0.0844***	0.0354*
	(0.0123)	(0.0121)	(0.0146)
Panel B: Income Quartile			
Second	0.0859***	0.1984***	0.0578**
	(0.0157)	(0.0155)	(0.0187)
Third	0.1972***	0.4925***	0.1358***
	(0.0167)	(0.0165)	(0.0196)
Fourth(Highest)	0.1947***	0.5528***	0.1411***
	(0.0168)	(0.0166)	(0.1958)
Observations	48661	48808	39359
Standard errors in parentheses			
* p<0.05	** p<0.01	*** p<0.001	

Table 5: Psychological Well-being and Personal Characteristics: Full Sample Ordered Probit Model,continued

Independent Variable	**General Happiness**	**Satisfaction with financial situation**	**Satisfaction with Job**
Panel C: Working Status	(1)	(2)	(3)
Unemployed	-0.297***	-0.411***	-0.174***
	(0.0226)	(0.023)	(0.0243)
Self-employed	0.0460**	0.0554***	0.292***
	(0.0169)	(0.0166)	(0.0198)
Retired	0.102***	0.164***	
	(0.0194)	(0.0192)	
Keep House	-0.01	0.0710***	
	(0.0171)	(0.0169)	
School	0.180***	0.119**	
	(0.0384)	(0.0379)	
Other	-0.359***	-0.491***	
	(0.0403)	(0.0414)	
Age	0.00294***	0.0147***	0.00979***
	(0.000387)	(0.000384)	(0.00053)
Age squared	0.0000363***	0.000151***	0.000117***
	(0.00000475)	(0.00000471)	(0.00000581)
Prestige Score	0.0031***	0.000151***	0.0112***
	(0.00047)	(0.000464)	(0.000549)
White	0.0423	0.1726***	0.0735**
	(0.0239)	(0.0138)	(0.0264)
Panel D: Marital Status			
Married	0.513***	0.188***	0.114***
	(0.0119)	(0.0116)	(0.0136)
Divorced	-0.0643***	-0.137***	-0.0321
	(0.0165)	(0.0162)	(0.0196)
Separated	-0.405***	-0.235***	-0.0119
	(0.0284)	(0.0287)	(0.0322)
Never married	-0.240***	0.0199	-0.113***
	(0.0155)	(0.0154)	(0.0173)
Standard errors in parentheses			
* p<0.05	** p<0.01	*** p<0.001	

6 Extension: Psychological Well-Being and Economic Activities

A large number of studies have found that people who constantly feel happy behave and make decisions fundamentally different from those who are not. We summarize the studies that investigate how happiness affects people's health, consumption activities, working behaviors, and investment behaviors. Due to the availability of the data, we show some evidence using ordinary least squares(OLS) regression and probit regression. The more rigorous studies should be done in the future work with better quality of data.

6.1 Physical and Mental Health

The research about how people's psychological well-being affects physical and mental health mainly falls into the following two parts: how positive psychological well-being affects health and how negative self-feeling affects health. The former have shown evidence that positive affects are associated with lower morbidity, lower level of symptoms and pain, and higher longevity among older community-dwelling individuals(Pressman and Cohen, 2005). Straume and Vittersø (2015) also found evidence that happiness is negatively related to sick-leave. Positive affectivity is a strong predictor of good physical health (Billings et al., 2000) but negative affectivity doesn't show evidence to predict health symptoms.(Joiner,2001).

On the other hand, people constantly in an unhappy state are found to be related to a higher level of stress and poor psychological health, as well as self-reported physical health(Watson and Pennebaker, 1989). The magnitude of the effects of negative affects is also related to the culture and industrialization level and democratization level(Curhan et al., 2014).

Our study sheds a light on detecting the state of people's psychological well-being by using individual traits and macroeconomic indicators. Making it possible for re-

Table 6: Happiness and Health

	(1) Health	(2) Suicide if incurable disease	(3) Suicide if bankrupt	(4) Suicide if dishonored family
Happiness	0.408***	-0.134***	-0.091***	-0.097***
	(0.012)	(0.014)	(0.019)	(0.019)
Individual Controls	Yes	Yes	Yes	Yes
Observations	35612	28295	28876	28824

Standard errors in parentheses
* p<0.1 ** p<.05 *** p<.01

	(5) Suicide if tired of living	(6) Ever test HIV	(7) Days of poor mental health
Happiness	-0.107***	-0.099***	-3.271***
	(0.016)	(0.021)	(0.129)
Individual Controls	Yes	Yes	Yes
Observations	28569	11217	8430

Standard errors in parentheses
* p<0.1 ** p<.05 *** p<.01

searchers and policymakers to be able to focus on those who tend to be mentally disadvantaged. Column 2-5 of table 6 shows the unhappy people are more likely to consider suicide when negative events happen in their lives. Happier people have better self-reported health status as shown in column (1). Happier people are less likely to have HIV test and have less days of poor mental feelings. With more data available, policymakers can use our models to target people who are in need of special help and make these studies more cost-efficient.

6.2 Consumption Activities

A growing amount of literature has been a focus on how happiness affects an individual's consumption choices. It has been proved in many literature people's emotions, such as happiness are important predictors in forecasting consumers' choices. A vast literature shows that people in a happy feeling consumes systematically different types of goods from those who are not. For example, people who feel happy are less likely to choose risky options(Isen and Patrick 1983), more likely to make healthier choices (e.g., less alcohol drinking and cigarette consumption). Isen(2001) found that happiness leads to helping and interpersonal understanding which implies increasing in customer satisfaction.

Unhappiness tends to motivate smoking behavior(Brandon,1994). Yet the relations between smoking and negative affect are more complicated. People who are stressed, angry and unhappy feelings reported smoking more(Coan, 1973; Becoña et al., 1999). The study also found more alcohol and marijuana use among unhappy college students(Magid, Colder, and et al.,2009). In table 7, we verify their findings.

Happiness is an important indicator but it is not easy to measure and observe. Many current studies fail to take it into consideration due to a lack of data. Our study provides a tool to classify the state of people's psychological well-being, which can be a potentially very important instrument for people in business and commercial to predict consumers' psychological mental state and make corresponding strategies.

Table 7: Happiness and Risky Behavior

	(1) Whether Smoking	(2) Ever quit smoking	(3) Ever drink	(4) Drink too much
Happiness	-0.182***	0.04	-0.097***	-0.068***
	(0.02)	(0.034)	(0.022)	(0.024)
Individual Controls	Yes	Yes	Yes	Yes
Observations	13561	4289	13563	9825

Standard errors in parentheses
* p<0.1 ** p<.05 *** p<.01

6.3 Working Behaviors

There is a large literature on productivity and personal happiness level(Siebert and Zubanov 2009). Edmans(2012) found that individuals' satisfaction with jobs is an important predictor of their stock market performance. Increasing job satisfaction increases value-added per hour working in manufacturing by 6.6%(Bckerman and Ilmakunnas, 2010).

Isen(2001) showed that happiness also affects doctor-patient interaction and medical decision making by increasing more understanding between doctors and patients. It is suggested that happiness should be considered in policy decisions as well.

In table 8, we show that happier people are likely to have more hours working and they are more likely to still work if they get rich.

Table 8: Happiness and Working

	(1) Hours worked per week	(2) Work if rich
Happiness	0.289**	0.049***
	(0.14)	(0.016)
Individual Controls	Yes	Yes
Observations	31211	22643
Standard errors in parentheses		
* p<0.1	** p<.05	*** p<.01

6.4 Investment Behaviors

Happier people have a different attitude to taking risks than less happy individuals. Labroo and Patrick(2009) have found that people with positive feelings are more likely to adopt for future goals while people in negative moods are more likely to focus on immediate and proximal events. Their finding has some important implications in consumers' investment decision. In economic literature, identifying economic agents' myopia is an important topic. Delis and Mylonidis (2015) found that happiness lowers the probability of investing in risky assets and insurance. Rao, Mei, and Zhu(2016) found evidence that happier people have higher stock market participation potentially due to more trust in capital and optimism. Chuluun and Graham (2015) found a positive correlation between local happiness and firm investment and R&D and the effect is larger for young firms.

The column (3) of 9 show that happier people are more confident in financial and banking systems. The investor confidence has long been proved to significantly affect their investment decisions and even the macroeconomic conditions. But few economic studies show the reason why some investors have more confidence. The reasons are

Table 9: Happiness and investment behavior

	(1) Whether Own stock	(2) Whether Own Option	(3) Confidence in Financial Institutes
Happiness	-0.023	0.013	0.180***
	(0.043)	(0.068)	(0.011)
Individual controls	Yes	Yes	Yes
Observations	3751	1699	33477

Standard errors in parentheses
* p<0.1 ** p<.05 *** p<.01

complicated and we show one aspect of that.

The state of happiness is crucial and should be an indicator that the financial manager would like to know. But it is not easy to obtain. Our findings help future research in classifying people's psychological well-being state and make the quantification possible.

7 Conclusion

This paper shows that using machine learning method and a set of crucial features, we can detect individuals with psychological disadvantages relatively precise. It also suggests a new way to analyze the determinants of happiness.

We use psychological well-being data in the United States. The data is formed of questions with subjective answers, such as *"How happy are you?", "How satisfied are you with your financial situation?"*, and *"How satisfied are you with your job?*. We also estimated the ordered probit model. We found that marriage matters. People who are married reported higher level of happiness. Income is also important for all three PWB measurements. Prestige of occupation is another crucial features in predicting happiness and satisfaction and it has been under covered in the current economics of happiness studies.

The impacts of unhappiness are large. It has negative impact on self-reported physical health and mental health. Unhappiness increases the probability of risky behaviors, such as smoking and excessive drinking. Happiness, on the other hand, increases people's confidence in financial institutions, productivity and affection towards their career.

In summery, this paper provides a comprehensive study of economic of psychological well-being. The methods developed in this paper have broad applications for economists who analyze the psychological impact on economic decisions and behaviors and policy-makers who seek to target psychological disadvantaged individuals.

Appendix

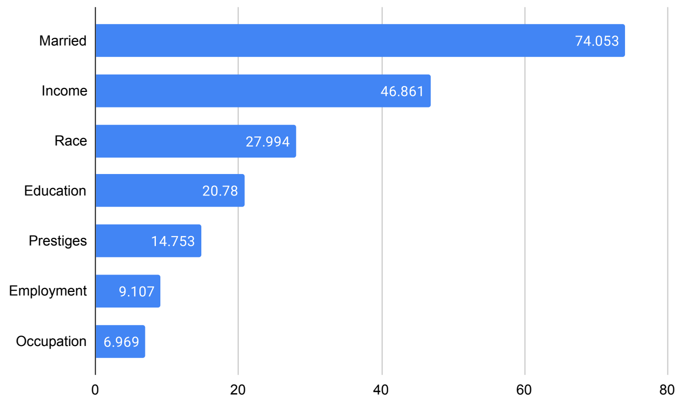
A: Feature description

Appendix A: Feature Description

Feature	**Type**	**Details**
Age	Numeric	Range between 18-89
Marital Status	Binary	Married or not
Income	Continuous	Adjusted to 2012 dollars
Race	Binary	White or others races
Education	Numeric	Highest years of schooling
Religion	Categorical	Types of Religion
Occupation	Categorical	Types of Occupation
Prestige Score	Numeric	Calculated via type of occupation
Children	Numeric	Number of Children
Working Status	Binary	Currently Employed or not

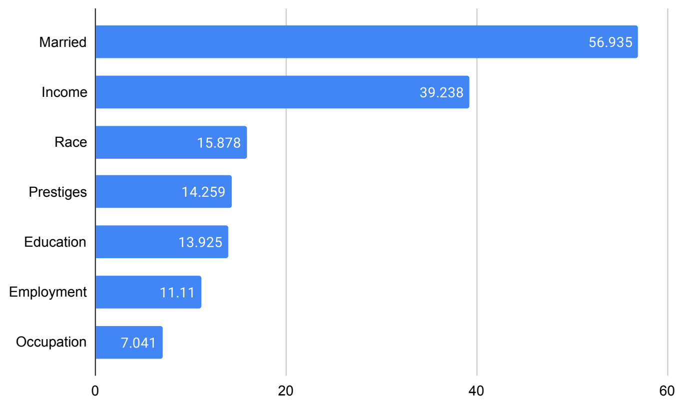
B: Importance Score

Figure 6: The Importance Score for General Happiness During Recession



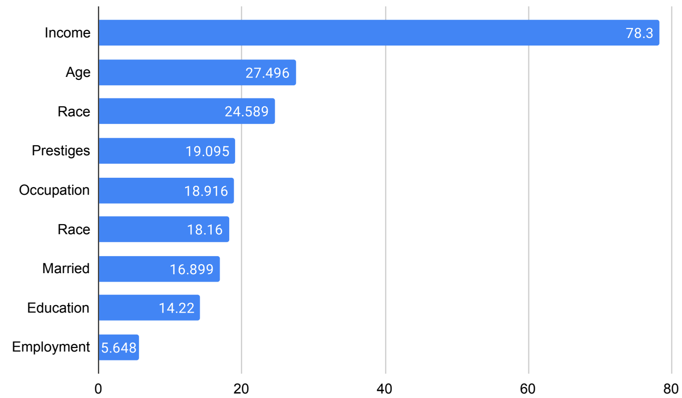
Note: the importance scores are scaled to 100.

Figure 7: The Importance Score for General Happiness During Expansion



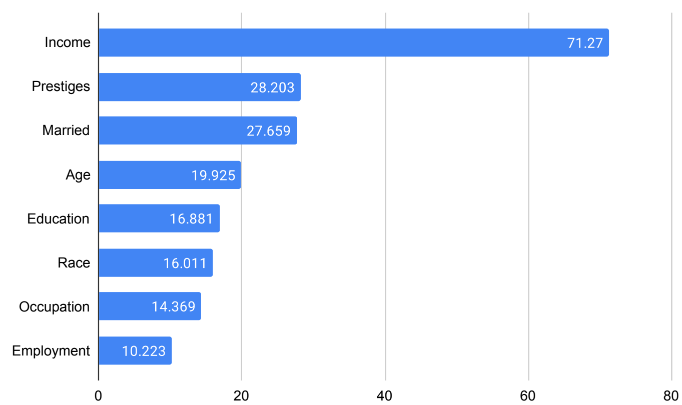
Note: the importance scores are scaled to 100.

Figure 8: The Importance Score for Satisfaction for Financial Situation During Recession



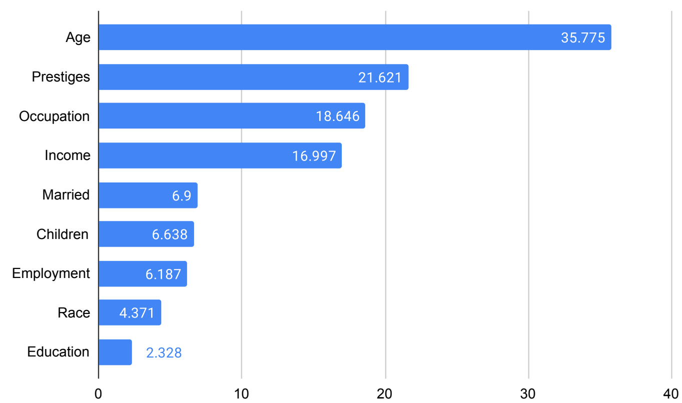
Note: the importance scores are scaled to 100.

Figure 9: The Importance Score for Satisfaction for Financial Situation During Expansion



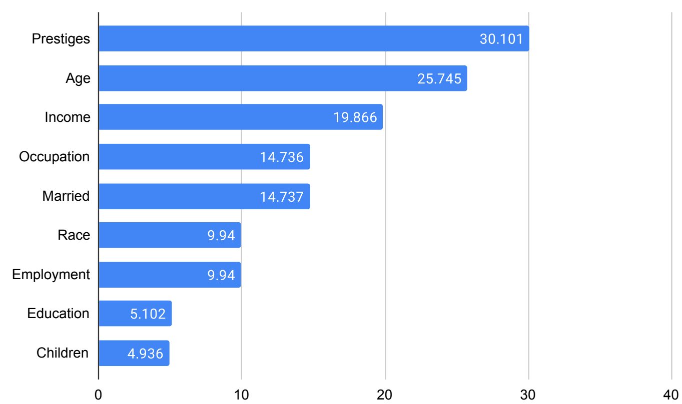
Note: the importance scores are scaled to 100.

Figure 10: The Importance Score for Satisfaction for Job During Recession



Note: the importance scores are scaled to 100.

Figure 11: The Importance Score for Satisfaction for Job During Expansion



Note: the importance scores are scaled to 100.

C: classification for general happiness with binary outcomes

Classification for General Happiness $k = 7$

	Classified		
True	Not Happy	Happy	Total
Not Happy	5132	884	6016
	85.31%	14.69%	100%
Happy	15695	27012	42707
	36.75%	63.25%	100%
Total	20827	27896	48723
	42.75%	57.25%	100%
Priors	0.5	0.5	

References

Becoña, E., Vásquez, F. L., Fuentes, M. J., and Lorenzo, M. d. C. (1999). Anxiety, affect, depression and cigarette consumption. *Personality and Individual Differences*, 26(1):113–119.

Billings, D. W., Folkman, S., Acree, M., and Moskowitz, J. T. (2000). Coping and physical health during caregiving: The roles of positive and negative affect.

Blanchflower, D. G. and Oswald, A. (2007). Is well-being u-shaped over the life cycle? Working Paper 12935, National Bureau of Economic Research.

Brandon, T. H. (1994). Negative Affect as Motivation to Smoke. *Current Directions in Psychological Science*, 3(2):33–37.

Bckerman, P. and Ilmakunnas, P. (2010). The job satisfaction-productivity nexus: A study using matched survey and register data. *Industrial and Labor Relations Review*, 65.

Cetre, S., Clark, A. E., and Senik, C. (2016). Happy People Have Children: Choice and Self-Selection into Parenthood. *European journal of population = Revue europeenne de demographie*, 32(3):445–473.

Chen, F., Hsu, C.-L., Lin, A. J., and Li, H. (2020). Holding risky financial assets and subjective wellbeing: Empirical evidence from china. *The North American Journal of Economics and Finance*, page 101142.

Chuluun, T. and Graham, C. (2015). Local Culture and Firm Behavior: Do Firms in Happy Places Invest More? *SSRN Electronic Journal*.

Clark, A. E. and Oswald, A. J. (1994). Unhappiness and Unemployment. *The Economic Journal*, 104(424):648–659.

Coan, R. W. (1973). Personality variables associated with cigarette smoking.

Curhan, K. B., Sims, T., Markus, H. R., Kitayama, S., Karasawa, M., Kawakami, N., Love, G. D., Coe, C. L., Miyamoto, Y., and Ryff, C. D. (2014). Just how bad negative affect is for your health depends on culture. *Psychological science*, 25(12):2277–2280.

Danner, D., Snowdon, D., and Friesen, W. (2001). Positive emotions in early life and longevity: Findings from the nun study. journal of personality and social psychology, 80(5), 804-813. *Journal of personality and social psychology*, 80:804–13.

Deaton, A. and Stone, A. A. (2016). Understanding context effects for a measure of life evaluation: how responses matter. *Oxford Economic Papers*, 68(4):861–870.

Delis, M. and Mylonidis, N. (2015). Trust, happiness, and households financial decisions. *Journal of Financial Stability*, 20(C):82–92.

Di Tella, R., MacCulloch, R. J., and Oswald, A. J. (2001). The Macroeconomics of Happiness. Technical report.

Easterlin, R. A. (1974). Does economic growth improve the human lot? some empirical evidence. In DAVID, P. A. and REDER, M. W., editors, *Nations and Households in Economic Growth*, pages 89 – 125. Academic Press.

Edmans, A. (2012). The link between job satisfaction and firm value, with implications for corporate social responsibility. *Academy of Management Perspectives*, 26(4):1–19.

Frey, B. S. and Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, 40(2):402–435.

Grn, C., Hauser, W., and Rhein, T. (2010). Is any job better than no job? life satisfaction and re-employment. *Journal of Labor Research*, 31(3):285–306.

Helliwell, J. F., Layard, R., and Sachs, J. (2012). World happiness report [2012].

Kapteyn, A., de Geer, S. V., de Stadt, H. V., and Wansbeek, T. (1997). Interdependent preferences: An econometric analysis. *Journal of Applied Econometrics*, 12(6):665–686.

Labroo, A., Patrick, V., served as editor, J. D., and served as associate editor for this article., M. F. L. (2009). Psychological distancing: Why happiness helps you see the big picture. *Journal of Consumer Research*, 35(5):800–809.

Lucas, R. E., Clark, A. E., Georgellis, Y., and Diener, E. (2003). Reexamining Adaptation and the Set Point Model of Happiness: Reactions to Changes in Marital Status.

Magid, V., Colder, C., Stroud, L., Nichter, M., Nichter, M., and Abrams, D. (2009). Negative affect, stress, and smoking in college students: unique associations independent of alcohol and marijuana use. *Addictive Behaviors*, 34(11):973–5.

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

Pressman, S. and Cohen, S. (2005). Does positive affect influence health? *Psychological bulletin*, 131:925–71.

Siebert, S. and Zubanov, N. (2009). Searching for the optimal level of employee turnover: A study of a large u.k. retail organization. *Academy of Management Journal*, 52:294–313.

Stevenson, B. and Wolfers, J. (2009). The paradox of declining female happiness. *American Economic Journal: Economic Policy*, 1(2):190–225.

Straume, L. V. and Vittersø, J. (2015). Well-Being at Work: Some Differences Between Life Satisfaction and Personal Growth as Predictors of Subjective Health and Sick-Leave. *Journal of Happiness Studies*, 16(1):149–168.

Walczak, B. and Massart, D. (2000). Chapter 15 - calibration in wavelet domain. In Walczak, B., editor, *Wavelets in Chemistry*, volume 22 of *Data Handling in Science and Technology*, pages 323 – 349. Elsevier.

Watson, D. and Pennebaker, J. W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity.

Winkelmann, L. and Winkelmann, R. (1998). Why are the unemployed so unhappy?evidence from panel data. *Economica*, 65(257):1–15.