

# Econometric Investigation of The Determinants of Well-being Through Empirical Analysis

Wanjun Xiao<sup>1, a</sup>

<sup>1</sup>The University of Sheffield, Western Bank, Sheffield S10 2TN, England

<sup>a</sup>pthahnil@163.com

**Abstract:** Utility functions serve as a prevalent basis for the theory of maximisation in the field of microeconomics. Consequently, conducting empirical research on life satisfaction offers a robust means of assessing the applicability of this theory. Through the utilisation of data derived from the UK Household Longitudinal Survey, an empirical examination may be undertaken to explore the factors influencing happiness, employing econometric techniques. This analysis aims to identify certain demographic groups that may require supplementary assistance. By employing linear and non-linear probability models, our analysis reveals that maintaining good health significantly contributes to the likelihood of experiencing contentment, but the influence of wealth on this probability may be less substantial than initially anticipated.

**Keywords:** Life satisfaction, Determinants of happiness, Probability models.

## 1. Introduction

In recent years, there has been concern that standard macroeconomic statistics such as gross domestic product, long used as a proxy for measuring well-being, fail to truly reflect people's current and future living conditions. The ongoing economic and financial crisis has reinforced this perception, and it is now widely recognized that GDP figures can only partially reflect a wide range of factors that are critical to people's lives. At the heart of policy action must be the needs, concerns and aspirations of the people and the sustainability of our society. The OECD publication "How's Life? 2020: Measuring Well-being" provides a good starting point for studying what went wrong. The OECD report categorizes itself into two categories: 1) positive and negative effects, which we can perceive as feelings of happiness, and 2) life satisfaction, which is an assessment of our well-being by reflecting on our current lives degree. Life satisfaction can be expressed by the happiness index. The happiness index is a topic of the times that the government is concerned about. If GDP (gross domestic product) and GNP (gross national product) are the standards for measuring the wealth of the country and the people, then the Gross National Happiness (GNH) is the standard for measuring people's happiness.

The field of happiness studies has evolved. Economic forces gave way to non-economic ones as influencing variables. According to classical economic thinking, increasing one's wealth is the most potent way to improve one's happiness, hence increasing one's wealth appears to entail increasing one's happiness. However, a psychology department research of the link between money and happiness has come up with some interesting results: The degree of link between people's income and happiness is lower when a country's income level is lower. The association is quite tight, but once this threshold is crossed, it lessens or even vanishes. Income level influences people's well-being by no more than 2% of the variables that affect their well-being. Easterlin's paradox, often known as the happiness gap, is a phenomena where happiness and income are in conflict (Happiness Gap). Various studies have shown that when income levels are very low, the relationship between income and happiness is

stronger. However, as economic development progresses, the effect of wealth on happiness gradually weakens, while other factors such as career achievement, educational level, marital quality, religious belief, life events, social support, and other factors gradually increase the influence of happiness. Modern research increasingly switches its focus from economic to non-economic elements, from univariate to multivariate, and emphasises multivariate's overall influence. The majority of past studies have been cross-sectional surveys and associated studies. Longitudinal design studies and cross-lagged survey designs are examples of this type of research. The majority of past studies have been cross-sectional survey research and associated studies. In order to understand the cause and impact of well-being, longitudinal design studies and cross-lagged survey designs are used to explore probable causal links between predictors and predicted variables.

The existing research is very comprehensive, but we will inevitably find some problems. For example, the standards of life satisfaction are not uniform, which leads to some conflicts between the conclusions of the studies. Our study will convert the life satisfaction variable into a binary variable equal to 1 if the respondent reports being mostly satisfied with their life and equal to 0. Find the relationship between them and life satisfaction by picking other variables, modeling and analyzing.

## 2. Literature Review

Extensive study has been conducted on the determinants of life satisfaction, given its significance for societal implications. For instance, Herzog and Rodgers (1981) conducted a study which revealed that older individuals, despite experiencing lower income as a result of retirement and poorer health indicators, generally expressed higher levels of satisfaction compared to younger and middle-aged individuals (Herzog and Rodgers, 1981, p.143). However, it is important to note that this finding may be influenced by factors such as a tendency to please others or adaptation effects (p.160).

Stutzer and Frey (2006) conducted a study to examine the relationship between happiness and marriage, specifically

investigating whether individuals who are already happy are more likely to get married, or if marriage itself leads to increased happiness. According to Stutzer and Frey (2006, p.328), their research indicates that those who are married tend to experience higher levels of satisfaction compared to those who are single or widowed. In their study, Kahneman and Deaton (2010) observed that when examining gender differences, women exhibited a twin-peaked utility pattern. Specifically, women reported higher instances of life satisfaction compared to males. However, it is important to note that women also had elevated degrees of stress and melancholy (p. 16491). According to Clark (2018), there is a potential correlation between better satisfaction ratings and expectations. The author highlights that women tend to have lower job market expectations, which results in a bigger boost in happiness when they secure employment (p.248). This phenomenon is similarly observed in the context of higher education, where pursuing further educational opportunities is expected to enhance employment prospects and future earnings, leading to heightened happiness. However, it is important to note that this pursuit also gives rise to elevated expectations.

Moreover, recent scholarly investigations have revealed that the acquisition of wealth does not inevitably lead to happiness. For instance, Kahneman and Deaton (2010) have identified a threshold of approximately \$75,000, beyond which the incremental increase in happiness diminishes significantly (p.16492). Furthermore, alternative studies propose that it is not the absolute income level that influences satisfaction, but rather one's relative socioeconomic standing. As an illustration, the Easterlin Paradox (Easterlin, 1974) discovered that the overall happiness of a nation would exhibit a positive correlation with money over a fixed timeframe, but not during a longitudinal study, due to people experiencing a decrease in satisfaction as their peers achieve higher levels of wealth.

Blanchflower and Clark (2020) have provided evidence indicating that the presence of negative coefficients of happiness about parenthood may be primarily ascribed to the financial strain experienced by parents who have challenges in meeting their financial obligations (p. 628).

In their study, Diener and Chan (2011) examined the relationship between health and satisfaction, specifically focusing on the impact of smoking on life satisfaction. Their findings indicated that those who did not smoke exhibited greater levels of life satisfaction (p. 13). Moreover, those who experience higher levels of happiness tend to have enhanced immune system functioning and a reduced susceptibility to cardiovascular ailments (p.12).

The subsequent section will employ these variables, together with some factors that have been developed independently on Stata, to examine the overall impacts on life satisfaction in the United Kingdom, commencing with an Ordinary Least Squares (OLS) regression.

### 3. Analysis of Data and Methodology

For a binary dependent variable of being satisfied, the probability of being satisfied for an OLS model is given by  $prob(y = 1|x) = F(x, \beta)$  and the probability of not being in debt is  $prob(y = 0|x) = 1 - F(x, \beta)$  where  $F(x, \beta) = \beta'x$ .

Since the expectation of  $y$  given  $x$  is  $F(x, \beta)$ , we can write the regression model as  $y = E[y|x] + (y - E[y|x]) = \beta'x + \varepsilon$ . Since our models include dummy variables, we

minimise the sum of squared residuals subject to the constant and dummy coefficients, such that  $\hat{\alpha} = \bar{Y} - \hat{\delta}\bar{D} = \frac{\bar{Y}_{D=0}}{\bar{D}=0}$  and  $\hat{\delta} = \frac{\sum(D_i - \bar{D})(Y_i - \bar{Y})}{\sum(D_i - \bar{D})^2} = \bar{Y}_{D=1} - \bar{Y}_{D=0}$

Table 1 shows a standard OLS regression on the binary outcome variable "satisfied", set equal to 1 when the individual reports being mostly or completely satisfied with their life. Most coefficients are significant at the 1% level, but the low R2 indicates a weak goodness of fit. For our continuous variables, the coefficients represent a percentage point (pp) change (divided by 100) in the probability of reporting being satisfied. A 1 year increase in age would, ceteris paribus, increase the probability of reporting being satisfied by  $(0.00311 \times 100) = 0.31$  pp, whereas a £1000 increase in gross household monthly income (as the income variable is divided by 1000) would result in a 1.49 pp increase. Alternatively, dummy variables represent the pp change when the dummy is set equal to 1, or in other words when the condition for the dummy is met. Therefore, if an individual is in a relationship, they would have a 4.6 pp increase in their probability of being satisfied, alternative to a 4.38 pp decrease for those no longer in a relationship. Having a degree increases the probability by 3.77 pp, and women have a 0.32 pp increase, but there is little evidence to support this due to the low t-statistic. It is worth modifying the regression slightly to aim for a better goodness of fit, Income transformed into its natural log to represent a more uniform percentage growth across incomes, rather than a lump-sum increase.

**Table 1. OLS regression**

	satisfied
age	0.00311*** (14.14)
kids	-0.0427*** (-5.78)
incouple	0.0460*** (5.18)
separated	-0.0438*** (-3.62)
degree	0.0377*** (6.17)
income	0.0149*** (13.98)
female	0.00324 (0.55)
_cons	0.306*** (24.74)
N	28222
R2	0.030

t statistics in parentheses

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

In Table 2, income has been transformed into its natural log to represent a uniform percentage growth. The log of income is statistically significant, and indicates that when all variables are included, a 1% increase in gross household monthly income will increase the probability of reporting being satisfied by 0.031 pp (note the absence of multiplying the coefficient by 100 as it is now an elasticity). The dummy variables for smoking and being in good health are both statistically significant at the 1% level and have conflicting effects, matching Diener and Chan's (2011) findings. Since there exists some overlap between the smoker and goodhealth variables, regressions from here shall proceed only with

goodhealth. It should be noted that the large  $R^2$  may indicate some unwanted collinearity between the measures for being in good health and being satisfied, as both are subjectively

reported through surveys rather than objective measures, but there is no direct evidence and a VIF score returned no indication of this.

**Table 2.** Comparison of results when goodhealth/smoker are omitted/included

	satisfied	satisfied	satisfied	satisfied
age	0.00420*** (19.80)	0.00283*** (12.79)	0.00441*** (20.84)	
kids	-0.0436*** (-6.21)	-0.0442*** (-6.00)	-0.0435*** (-6.17)	
incouple	0.0293*** (3.43)	0.0392*** (4.39)	0.0306*** (3.59)	
separated	-0.0234** (-2.03)	-0.0378*** (-3.14)	-0.0261** (-2.27)	
degree	0.00543 (0.93)	0.0310*** (5.09)	0.0112* (1.92)	
lincome	0.0311*** (7.62)	0.0492*** (11.53)	0.0341*** (8.35)	
female	0.00571 (1.01)	-0.000450 (-0.08)	0.00795 (1.41)	
smoker	-0.0837*** (-9.81)	-0.130*** (-14.60)		-0.122*** (-14.37)
goodhealth	0.375*** (53.14)		0.382*** (54.33)	0.348*** (50.06)
_cons	-0.203*** (-5.62)	0.0102 (0.27)	-0.257*** (-7.21)	0.297*** (46.73)
N	28222	28222	28222	28222
R2	0.124	0.036	0.121	0.093

t statistics in parentheses

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

The age variable holds little explanatory power despite its significance, possibly as the age observations ranging from 16 to 103 in the dataset. Therefore, dummy variables for elderly (60+) and middle-aged (40-59) are generated to provide more insight and estimated against the baseline category of younger individuals to avoid the dummy variable trap. Another dummy variable is added to demonstrate the effects from whether the individual belongs to an ethnic minority group or not.

The probit model is formed by constructing latent variables, given by  $y^* = x'\beta + \varepsilon$  where  $y = 1$  if  $y^* > 0$ ;  $y = 0$  if  $y^* \leq 0$  and the error term is i.i.d standard normally distributed, with the probability of observing a positive outcome:

$$\Pr(y = 1|x) = \Pr(x'\beta + \varepsilon > 0) = \Pr(\varepsilon > -x'\beta) \quad (1)$$

$= 1 - G(-x'\beta) = G(x'\beta)$  and due to the error specification, we are left with:

$$G(\alpha + \beta X_i) = \Phi(\alpha + \beta X_i) \equiv \int_{-\infty}^{\alpha + \beta X_i} \phi(v) dv \quad (2)$$

When using the probit command in Stata, it calculates

coefficients through the log-likelihood function:  $LL = \sum_{i=1}^n \ln G(q_i x'_i \beta)$  where  $(q_i = 2y_i - 1)$  and from this we calculate the Marginal Effects (MEs) used for interpretation via the following function:

$$\frac{\delta \Pr(y=1|x)}{\delta x_j} = g(x'\beta) \frac{\delta E(y^*|x)}{\delta x_j} = g(x'\beta) \beta_j \quad (3)$$

Table 3 shows both OLS and probit results. Splitting up the age observations into separate categories has not caused any problems in Table 3, with the elderly having a 17.6 pp increase in the probability of being satisfied to those under 40. We also observe that minorities have a 7.04 pp decrease in the probability of being satisfied, whereas having between 7-9 hours of sleep per night increases the probability by 10.2 pp. Table 3 shows only minor changes to the coefficients/MEs and standard errors, indicating the LPM may provide sufficient explanation. However, the predicted probability of being satisfied for an individual in the baseline category for all variables (denoted by the constant) would be negative, which is not possible. Furthermore, the variance of the error term depends on the value of the parameter, leading to inefficient standard errors due to heteroskedasticity. Therefore, the NLPM shall be employed from this point as it

will not impede upon any potential non-linear findings.

In tables one to three, having children has produced a negative impact on the probability of reporting being satisfied. Table 4 demonstrates the observed changes for when we introduce a dummy variable for having bills paid into the model, as well as its interaction term with having children. The inclusion of these variables converts the children variable from -0.0252 in Table 3 to 0.0371 here, indicating that when payment of bills is held constant, having children increases the probability of reporting being satisfied by 3.71 pp, however this is insignificant at the 10% level. Additionally, the interaction effect is negative, suggesting that a parent with

no outstanding bills would be less likely to be satisfied than a parent with outstanding bills – this seems inconsistent as debt should not provide satisfaction, but it is possible there are unobserved exogenous factors biasing this estimate, such as expenditure on consumption goods instead of bills. Furthermore, Table 5 represents the ME of having children for parents in a relationship compared to those who are single. The results contradict the findings of Blanchflower and Clark (2020), predicting that parents in a relationship are less likely to report being satisfied than single parents, with the difference equating to 5.08 pp.

**Table 3.** OLS and Probit regressions with minority and goodsleep parameters on the dependent variable of “Satisfied”

	OLS	Probit
1.elder	0.174*** (0.00903)	0.176*** (0.00856)
1.middleaged	-0.0163** (0.00759)	-0.0170** (0.00753)
1.kids	-0.0261*** (0.00697)	-0.0252*** (0.00690)
1.incouple	0.0735*** (0.00810)	0.0723*** (0.00814)
1.separated	0.0232** (0.0109)	0.0195* (0.0108)
1.degree	0.0185*** (0.00576)	0.0190*** (0.00575)
1income	0.0346*** (0.00402)	0.0342*** (0.00398)
1.goodhealth	0.355*** (0.00703)	0.357*** (0.00671)
1.minority	-0.0724*** (0.00796)	-0.0704*** (0.00799)
1.goodsleep	0.102*** (0.00565)	0.102*** (0.00571)
1.female	0.00978* (0.00558)	0.00935* (0.00557)
_cons	-0.162*** (0.0332)	
N	28222	28222
R2	0.143	

Standard errors in parentheses  
\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

**Table 4.** Probit marginal effects for kids and bills

	Satisfied
Age	0.00431*** (0.000208)
Has children	0.0371 (0.0260)
In relationship	0.0270*** (0.00845)
No longer in relationship	-0.0243** (0.0115)
Up to date with all bills	0.212*** (0.0188)
Has kids and bills paid	-0.0737*** (0.0272)
Log of monthly income	0.0281*** (0.00397)
In good health	0.356*** (0.00677)
Individual has recommended amount of sleep per night	0.108*** (0.00572)
Is female	0.00972* (0.00558)
Observations	28222

Standard errors in parentheses  
\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

**Table 5.** Marginal effect of kids on satisfaction for couples against non-couples, with difference listed below

	satisfied
1.kids 0.incouple	-0.00890 (-0.73)
1.incouple	-0.0597*** (-7.61)
N	28222

z statistics in parentheses  
\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Binary dependent variable models are useful for demonstrating the overall effect that a parameter will have, but sometimes it is useful to analyse the effects observed for specific demographics. In this instance, the use of ordered outcome variables is necessary. For life satisfaction there are seven possible response categories, ranking from the lowest response (1) being completely dissatisfied to the highest (7) being completely satisfied. Thus far we have considered “satisfaction” as responding with a score of six or seven, but now we may examine the categories in more detail. Table 6 shows the ME of being in good health on the probability of responding within a certain satisfaction category. An

individual in good health is 5.22 pp less likely to respond being “completely dissatisfied” with life than someone not in good health, whereas they would be 21.9 pp more likely to report being “mostly satisfied” with life.

The results from Table 6 were from an ordered outcome variable regression with several other parameters included, which can be found in Table 7. Although it is possible to compute every response category simultaneously for each variable, this makes the table relatively difficult to interpret. Thus, we select three categories in particular: “completely dissatisfied”, “neither satisfied nor dissatisfied”, and “completely satisfied”, to provide interpretations for the most extreme and midpoint responses.

**Table 6.** Ordered probit for good health on life satisfaction

	Life satisfaction response category
Individual in good health	
Completely dissatisfied	-0.0522*** (-23.73)
Mostly dissatisfied	-0.0743*** (-30.03)
Somewhat dissatisfied	-0.0974*** (-38.58)
Neither satisfied nor dissatisfied	
Somewhat satisfied	-0.0289*** (-23.86)
Mostly satisfied	0.219*** (47.09)
Completely satisfied	0.102*** (52.96)
Observations	28222

z statistics in parentheses  
\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

**Table 7.** Ordered probit for the lowest, middle, and highest response categories of life satisfaction

	Completely dissatisfied	Neither satisfied nor dissatisfied	Completely satisfied
Aged 60 or older	-0.0164*** (-19.83)	-0.0367*** (-23.85)	0.0799*** (22.64)
Aged 40-59	0.00224*** (3.15)	0.00438*** (3.19)	-0.00871*** (-3.22)
In relationship	-0.00600*** (-8.86)	-0.0116*** (-9.37)	0.0226*** (9.68)
Has a degree	0.000211 (0.37)	0.000415 (0.37)	-0.000836 (-0.37)
Log of monthly income	-0.00313*** (-7.70)	-0.00615*** (-7.92)	0.0124*** (7.94)
In good health	-0.0522*** (-23.73)	-0.0680*** (-42.01)	0.102*** (52.96)
Is a minority	0.00484*** (5.46)	0.00889*** (5.93)	-0.0172*** (-6.18)
Is female	-0.00124** (-2.25)	-0.00243** (-2.27)	0.00488** (2.27)
Individual has recommended amount of sleep per night	-0.0106*** (-16.10)	-0.0215*** (-18.11)	0.0411*** (19.36)
Observations	28222	28222	28222

z statistics in parentheses  
\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Table 7 provides more detailed analysis of previously explored parameters. Those aged 60 or older are 1.64 and 3.67 pp less likely to report being in categories one or four

respectively than people under the age of 40, whereas the middle aged are 0.22 and 0.43 pp more likely to report in these categories (the reverse effects hold true for the highest

category). Having a degree decreases the probability of being completely satisfied by 0.084 pp, however the ME for having a degree is insignificant at the 10% level. This could be due to the single indices assumption of the ordered response model, whereby relative magnitudes of the Marginal Probability Effects cannot vary across outcomes and individuals, such that:

$$\frac{MPE_{ijl}}{MPE_{ijm}} = \frac{\beta_l}{\beta_m} \quad (j \text{ denoting ordinal response categories}) \quad (4)$$

Furthermore, the single crossing property restricts MPEs to being positive at the lowest categories and negative at the highest or the opposite, not allowing for a bell or u-shaped distribution. A solution to this would be using a generalised ordered model, but this model requires much more computation for improved flexibility.

#### 4. Conclusion

Life happiness, as may be predicted, is a broad issue that cannot be fully described by a small set of characteristics; its maximum  $R^2$  value is around 14%. However, our probability models have shown substantial impacts from certain factors, including goodsleep and goodhealth. The research was shown to have limitations due to the absence of additional, more objective health metrics, such as the degree of physical activity. Significant estimations for various parameters were derived using ordinary least squares (OLS), probit, and ordered probit probability models. This outcome presents an intriguing opportunity to replicate the experiment using logit or generalised ordered estimation models.

#### References

- [1] Blanchflower, D. and Clark, A., 2020. Children, unhappiness and family finances. *Journal of Population Economics*, 34(2), pp.625-653.
- [2] Clark, A., 2018. Four Decades of the Economics of Happiness: Where Next?. *Review of Income and Wealth*, 64(2), pp.245-269.
- [3] Clark, A.E., and Blanchflower, D.G., (2021). "Children, unhappiness and family finances", *Journal of Population Economics*, 34, pp. 625-653
- [4] Diener, E. and Chan, M., 2011. Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity. *Applied Psychology: Health and Well-Being*, 3(1), pp.1-43.
- [5] EASTERLIN, R., 1974. Does Economic Growth Improve the Human Lot? Some Empirical Evidence. *Nations and Households in Economic Growth*, pp.89-125.
- [6] Herzog, A. and Rodgers, W., 1981. Age and Satisfaction. *Research on Aging*, 3(2), pp.142-165.
- [7] Kahneman, D. and Deaton, A., 2010. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38), pp.16489-16493.
- [8] Miao Yuanjiang. From happiness to happiness index—a research on happiness in development. *Nanjing Social Sciences*, 2009(11): 103-108.
- [9] OECD (2020), *How's Life? 2020: Measuring Well-being*, OECD Publishing, Paris, <https://doi.org/10.1787/9870c393-en>.
- [10] Stutzer, A. and Frey, B., 2003. Does Marriage Make People Happy, or Do Happy People Get Married?. *SSRN Electronic Journal*, 35(2).