An Investigation into the Factors that can Affect Life Satisfaction

Introduction

There can be no area of study more necessary, encompassing or relevant than the study of life satisfaction; as the principle factor in our lives, it forms the basis of every decision we make. Studying happiness does not come without difficulty; the hugely subjective nature of any data available and the many constituent variables that are unquantifiable and unsurveyable mean there can be no perfect model to predict an individual's satisfaction with their life. Despite this, there has been extensive research into the subject. Peterson et al. [2005] provide a summary of the history of hedonism and eudemonia, and seek to unify the two variables with another; the pursuit of engagement, as conceptualized by Seligman [2002]. Diener and Seligman [2002] contend that scrutiny of the happiest people among us will proffer answers as to how to improve the happiness of the rest of the population. Improving the lives of citizens must be the primary objective of any government, yet the simple measure of happiness is often overlooked, as economic prosperity and security take precedence. Whilst the holistic objectives of societal leaders may be a question for philosophers, we can utilize economics to evaluate data and the suitability of policy. This project will create and evaluate simple models analysing life satisfaction, focusing on the comparative difference in effect and significance of each variable.

Literature Review

Diener [1984] provides a comprehensive synopsis of past literature regarding life satisfaction, referencing a range of studies that find significant variables. One of the variables with

abundant previous literature is age; the relationship between age and happiness has been studied extensively. Many academics suggest a U-shaped relationship, yet Frijters and Beatton [2012] find a reverse-causality issue; happiness-increasing variables such as high income and marriage tend to occur at an age when individuals are already happy. Using logit models, they find a roughly wave-shaped relationship where happiness peaks at around 67, where life generally becomes more stress free, and deteriorates rapidly after 75, where health deteriorations often start. In this project both the raw age variable and an age^2 variable were analysed, as a form of sensitivity analysis to determine which was more suitable for the overall model. The age^2 variable showed the more significant result, so this was taken forward to the linear regressions

Argyle [1999] is one of many academics to have found a positive, albeit weak relationship between income and happiness. Diener et al (1985) studied some of the wealthiest U.S. citizens and found there was very little difference in their satisfaction levels when compared to the average U.S. individual. In this project, the natural logarithm of total monthly personal income will be used.

Income and education are intuitively closely linked; a positive correlation should logically be found if a regression were run. Studies have also found a direct link between education and happiness, indicating a relationship stronger than just the indirect effect from the income increase. The link between education and life satisfaction is not conclusive, but the majority of studies find a positive effect of education on happiness [Castriota, 2006].

The effect marital status has on life satisfaction is another area that has been researched extensively. This is one of the variables that will be focused on in the discussion segment of this project, as it is an area the Government can have a strong influence on. The current Marriage Allowance scheme in the U.K. allows the transfer of up to £1100 from a higher earning individual to their spouse, providing each meet income criteria, which could decrease their tax bills by up to £220 [Gov.uk, 2017]. The reasoning behind this is that incentivizing marriage financially should help to create more stable families.

Stack and Eshleman [1998], using an ordinary least squares regression model with data from 17 industrialized nations, found a significant positive correlation between life satisfaction levels and marital status, which was consistent across the nations. The authors cited intermediary stages as contributors to this relationship; marriage typically increases financial satisfaction and health satisfaction, which, in turn, increase happiness. However, they also accept that a major limitation of their study is the potential for social selection [Mastekaasa, 1992]; happier, healthier people are more likely to get married in the first place, creating issues of causality in their analysis. The use of a single-item, fixed response measure is a further limitation of their analysis, as multiple item measures have been found to have greater reliability [Diener, 1984]. Unfortunately, the data set used in this project also uses a single item measure for life satisfaction; this is unavoidable, yet is a limitation that must be taken into consideration.

There is debate as to whether a link exists between gender and happiness. The Campbell et al. [1976] study is one of a range to find little difference between the genders, yet it will be used as a control variable in this project for the sake of completeness. Another variable that

will be used in this project is the binary dummy variable of volunteering. Borgonovi [2008] analyzed the effect volunteering frequency has on individual's health and happiness levels; a significant relationship was found, with the paper stipulating the possibility that volunteering mitigates the tendency to care excessively about relative income. Again, however, this study uses a single item measure, which limits its reliability.

The effect of rurality will also be scrutinized as it has been linked to increased life satisfaction; Brown and Xu [2002], using probit models, found rural residents tended to have significantly higher levels of life satisfaction than urban residents, ceteris paribus. In a 2006 biomedical study, Groenewegen et al. show that green spaces in urban areas provide health and happiness benefits to the public. Rurality is another area that will be focused on in this project as the introduction of green spaces is a policy that could be implemented by the Government to the benefit of the public. Due to the intuitively strong relationship between health and happiness, health satisfaction will be used as a variable too.

Data Exploration

To begin, simple means and modes were analysed and scatter plots created to allow an overview of the data. One of these graphs can be seen by running the code provided; a scatter plot comparing the age and personal pay of the interviewees. Unsurprisingly, the graph shows an upward trend between the ages of 20 and 50, though after this the pattern is less clear. After removing NAs (in the data these were negative figures as seen in the graph, and after the filtering out of those earning above £14,500 per month, the curve can be seen more

clearly, and there appears to be a decline in average monthly pay from around the age of 60; possibly due to individuals retiring early, or working less as they age.
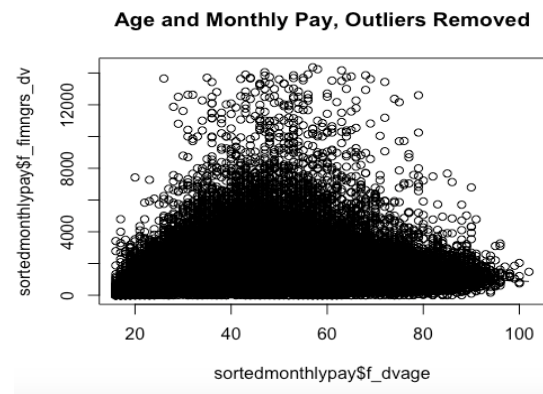
The 'y' variable also had to be picked at this early stage; in the data, four life satisfaction variables are present. This is because interviewees were asked to rank their subjective life satisfaction on a scale from 1-7, with 7 being the highest. The variable 'f_sclfsato' is the raw variable, 'Satisfied', 'SatisfiedN' and 'SatisfiedNN' are all binary variables taking different cut-off points for 'Satisfied' or 'Not Satisfied'. Ultimately, 'SatisfiedN' was chosen, which grouped responses 4-7 as 'Satisfied' and 1-3 as 'Dissatisfied'.

The table below shows how participants responded to the original question:

Response	Frequency	Valid Percent
Completely dissatisfied	691	1.9
Mostly dissatisfied	1916	5.4
Somewhat dissatisfied	2636	7.4
Neither satisfied nor dissatisfied	3171	8.9
Somewhat satisfied	6291	17.6
Mostly satisfied	16808	46.9
Completely satisfied	4294	12.0
Total	35807	100

It can be seen from observation that the modal response was 'mostly satisfied', whilst the mean value is 5.24. It can also be observed that even with the middle value included, responses 1, 2, 3 and 4 cumulatively have a lower frequency than responses 5, 6 and 7. This could be a reason to use the 'Satisfied' variable rather than the 'SatisfiedN' variable, as it

provides a more even split. However, responses 6 and 7 alone have higher frequencies than the rest combined, so 'SatisfiedNN' provides the most even split of the three.

<u>Methodology</u>

Before the model-building could begin, the data was cleaned to remove NAs and test and training sets were created, the test set being one-tenth of the size of the training set. These sets also excluded any variable that was irrelevant to this project; mainly those that were thought to have no substantial effect on life satisfaction.

To begin with, linear and logistic regressions were run, with the aim of confirming the expected effects of the variables. Once the significance of each variable was verified, each individual variable was analysed for its predictive power on life satisfaction using an RMSE model. All of the variables, both those in the final models and those that were only used for initial analysis, can be seen in the table below, along with the expected coefficients:

Variable	Description	Recoding	Expected Sign of Coefficient
Satisfied	Life satisfaction (neither satisfied nor dissatisfied=0)	Satisfied = 1 Dissatisfied = 0	N/A
SatisfiedN	Life satisfaction (neither satisfied nor dissatisfied=1)	Satisfied = 1 Dissatisfied = 0	N/A
SatisfiedNN	Life satisfaction (neither satisfied nor dissatisfied =0, mostly satisfied = 0)	Satisfied = 1 Dissatisfied = 0*	N/A
Age	Age	N/A	?
Age2	Age^2	N/A	?
Female	Binary dummy of gender	Female = 1 Male = 0	?
Unemployed	Whether an individual is unemployed or not	Unemployed = 1 Not Unemployed = 0	-ve
Employed	Self-employed and employees grouped together	Employee of self-employee = 1 Not employee or self-employee = 0	+ve
Student	Whether an individual is a full-time student or not	Student = 1 Not a student = 0	?
Volfreq0	Whether an individual had volunteered in the last 12 months	Volunteer = 1 Non-volunteer = 0	+ve
Rural	Whether an individual lives in a rural area or an urban area	Rural = 1 Urban = 0	+ve
Married	'Married', 'living as a couple' and 'in a registered same-sex civil partnership' grouped together	Married = 1 Non-married = 0	+ve
Degree	Whether an individual has a degree or not	Degree = 1 No degree = 0	+ve
Alevel	Individual's highest qualification is A level	A level = 1 Not A level = 0	+ve
GCSE	Individual's highest qualification is GCSE	GCSE = 1 Not GCSE = 0	+ve
Otherqual	Individual's highest qualification is another qualification	Other qualification = 1 Not other qualification = 0	?
lnperspay	Natural log of an individual's personal pay	N/A	+ve
Healthsat	Individual's satisfaction with their level of health	Satisfied = 1 Dissatisfied = 0	+ve

The table shows that the 'personal pay' variable was been transformed into logarithmic form; now called 'lnperspay'. This was to compensate for what is a vast range of responses, and should reduce the effect of outliers. The 'Other Qualification' variable has been left out to avoid issues of perfect multicollinearity; in the end, only 'degree' was used regardless. Health satisfaction is included as a binary variable, constructed in a similar way to life satisfaction; responses 1-4 were counted as 'unsatisfied' and 5-7 as 'satisfied'.

As the dependent variable is binary, an OLS linear regression model is unsuitable, as the possibility exists of individuals having predicted probabilities less than zero or greater than one. However, an OLS regression will still be run as a robustness check for other models. Logit and probit models were also be used, as cumulative distribution functions with sigmoid shapes are more suited to this type of binary data, though the results of these were not markedly different from the linear regression. The initial OLS model was as follows:

$$Z_i = \beta_0 + \beta_1 AGE_i + \beta_2 AGE2 + \beta_3 FEMALE_i + \beta_4 UNEMPLOYED_i + \beta_5 EMPLOYED_i + \beta_6 STUDENT_i + \beta_7 VOLFREQ0_i + \beta_8 RURAL_i + \beta_9 MARRIED_i + \beta_{10} DEGREE_i + \beta_{11} ALEVEL_i + \beta_{12} GCSE_i + \beta_{13} OTHERQUAL_i + \beta_{14} LNPERSPAY + \beta_{15} HEALTHSAT + \mu_i$$

Where Z_i = SatisfiedN$_i$

Next, models were run to analyse the residual mean squared errors of each variable compared to the 'y' variable. These models each included only the 'SatisfiedN' variable and the variable under scrutiny; a table was created to compare the RMSEs of each.

Empirical Results and Discussion

Initial runs of the linear OLS regression model were undertaken using 'Satisfied', 'SatisfiedN' and 'SatisfiedNN' as the dependent variables. It was clear from these regressions that 'SatisfiedNN' was the more accurate model; it had a higher R^2, fewer insignificant results and, as stated earlier, a more even split between satisfied and unsatisfied. This presents an issue; 'SatisfiedNN' groups 'somewhat satisfied' in the 'dissatisfied' category, which is clearly counter-intuitive. In order to produce the most accurate and intuitive set of results, 'SatisfiedN' was used, as a compromise. It should be made clear that the '0' category for 'SatisfiedNN' cannot be described as 'dissatisfied', as some in the 0 category may fall under 'somewhat satisfied'.

The results of the final model of the OLS regression can be seen below:

```
Call:
lm(formula = SatisfiedN ~ female + Employed + selfemploy + unemployed +
    retired + student + degree + married + volfreq0 + Age2 +
    Rural + lnperspay, data = trainset)

Residuals:
    Min      1Q  Median      3Q     Max
-0.98381 0.08887 0.12005 0.15781 0.35389

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.000e-01  1.577e-02  44.379  < 2e-16 ***
female       1.348e-02  4.371e-03   3.084  0.00205 **
Employed     1.329e-01  8.196e-03  16.211  < 2e-16 ***
selfemploy  -2.464e-02  8.125e-03  -3.032  0.00243 **
unemployed  -1.397e-02  1.355e-02  -1.030  0.30284
retired      1.702e-01  1.039e-02  16.384  < 2e-16 ***
student      1.776e-01  1.350e-02  13.156  < 2e-16 ***
degree       2.115e-02  4.596e-03   4.602 4.21e-06 ***
married      5.934e-02  4.664e-03  12.724  < 2e-16 ***
volfreq0     2.462e-02  5.281e-03   4.662 3.15e-06 ***
Age2        -4.256e-06  1.947e-06  -2.186  0.02884 *
Rural        1.514e-02  4.869e-03   3.109  0.00188 **
lnperspay   -3.194e-03  1.961e-03  -1.628  0.10347
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3473 on 27124 degrees of freedom
Multiple R-squared:  0.02936,   Adjusted R-squared:  0.02893
F-statistic: 68.38 on 12 and 27124 DF,  p-value: < 2.2e-16
```

As this model was a precursor to the RMSE model later, the individual figures will not be analysed in detail here, but it is interesting to note that nearly all of the variables showed significance. The R^2 is low at 0.03, but this was expected with such low coefficients; these variables do not impact life satisfaction majorly, but do have significance for the effect they have. Other regressions were run that included health satisfaction and leisure time satisfaction; these increased the R^2 to around 0.3; again expected, as satisfaction with a major section of life such as health will likely have a far bigger impact on life satisfaction that your gender or age.

The only result that did not provide significance where it was expected was the unemployed variable, though the sign was negative as expected. Due to possible issues with perfect multicollinearity, the regression was run again without this variable, though the results were almost identical.

Each significant variable was then used to calculate an RMSE; the three most interesting are included in the final code. These three can be seen in the table below:

	volmodelrmse	studentmodelrmse	ruralmodelrmse
1	0.8325269	0.906409	0.8128494

It can be seen from this table that the better explainer of life satisfaction is the rural variable; whether the individual lives in an urban or rural environment. Whether an individual volunteers or not is not far behind. Being a student has less impact than either; unsurprising,

especially given that the category of not being a student includes both employed and unemployed individuals, as well as retirees, and each of these has its own effect, as evidence in the linear regression earlier.

It is important to mention potential limitations of this project. One of the main problems faced, as touched upon earlier, is the simplistic nature of the dependent variable source question. Using just one question to determine life satisfaction does not allow sufficient breadth of analysis; the problems of using single-item measures was discussed in the literature review section of this project. Whilst this does not negate the validity of the analysis, more complete and accurate results could be drawn from multiple-item measures. The data under scrutiny was taken from a condensed version of 'Understanding Society' dataset; whilst this is a large data set with a sufficient sample size, it does lack certain variables that could have been utilized to improve the models. For example, the inclusion of religious beliefs or international data that includes several countries and cultures would enable a broader understanding of the economics of life satisfaction. Additionally, a higher R^2 could probably have been manufactured had certain other variables been included, such as job satisfaction and leisure time satisfaction. These were not included because of the potential for causal issues.

Conclusion

This project has gone some way to substantiating previous literature on many of the issues under scrutiny. The conclusions made here broadly agree with conclusions drawn from previous literature, with any minor discrepancies probably being due to miss-specification or

a lack or variation in the samples. The comparison of RMSEs appears to suggest that both

living in a rural environment (as opposed to in a city) has some significant, if small, effect on

life satisfaction, as does the act of volunteering, though to a lesser extent.

Bibliography

Argyle, M (1999) 'Causes and Correlates of Happiness', In: Kahneman D, Diener E, Schwarz N, editors. *Well-being: The foundations of hedonic psychology*, pp 353-373

Bolund and Hunhammar (1999) 'Ecosystem services in urban areas', *Ecological Economics* 29, pp 293-301

Borgonovi, F. (2008) 'Doing well by doing good. The relationship between formal volunteering and self-reported health and happiness', *Social Science & Medicine*, Volume 66 issue 11

Brown and Xu (2002) 'Separate family and community realities? An urban-rural comparison of the association between family life satisfaction', *Community, Work and Family*, Volume 5, No. 2

Campbell et al. (1976) 'The Quality of American Life' Available at: https://www.russellsage.org/publications/quality-american-life (Accessed: 12/03/2017)

Castriota, S (2006) 'Education and Happiness: a Further Explanation to the Easterlin Paradox?', *Department of Economics, Universita' Tor Vergata*, 302

Diener, E. (1984) 'Subjective well-being'. *Psychological Bulletin*, 95, pp 542-575

Diener et al. (1985) 'The Satisfaction with Life Scale' *Journal of Personality Assessment,* 49, 1

Diener and Seligman (2002) 'Very happy people'. *Psychological Science* , Volume 13, Issue 1, pp 81-84

Frijters and Beatton (2012) 'The mystery of the U-shaped relationship between happiness and age', *Journal of Economic Behaviour and Organisation*, Volume 82, Issue 2-3, pp 525-542

Gov.uk (2017) 'Marriage Allowance' Available at: https://www.gov.uk/marriage-allowance (Accessed: 10/03/2017)

Groenewegen et al. (2006) 'Vitamin G: effects of green space on health, well-being, and social safety', *BMC Public Health*, 6-149

Mastekaasa, A. (1992) 'Marriage and Psychological Well-Being: Some Evidence on Selection into Marriage', *Journal of Marriage and Family*, 54(4)

Peterson et al. (2005) 'ORIENTATIONS TO HAPPINESS AND LIFE SATISFACTION: THE FULL LIFE VERSUS THE EMPTY LIFE' *Journal of Happiness Studies*, 6:25–41

Seligman, M. (2002) 'Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfilment'. New York: Free Press.

Stack and Eshleman (1998) 'Marital Status and Happiness: a 17 Nation Study', *Journal of Marriage and Family*, Volume 60, Issue 2, pp 527-535

Note: The dataset used for this project is from https://www.understandingsociety.ac.uk Permission has been given for the use of this dataset in educational contexts such as this report.