Are you happier in town or the country?[1]

A quantitative analysis of the association between local authority mean estimates for self-reported happiness in England and Scotland and the proportion of local authority population resident in rural areas.

Lucky Dube

ABSTRACT

This project investigates the bivariate relationship between local authority mean estimates for self-reported happiness in England and Scotland and the proportion of local authority population resident in rural areas. Data are drawn from the 2011/12 Office for National Statistics headline estimates for personal wellbeing, and DEFRA and Scottish Government rural-urban population data. The results find a statistically significant association between the two variables. Analysis of subsets of data relating to England and Scotland find the association is stronger in Scotland. The project concludes by suggesting areas for further study and how they can support the development of public policy.

Programmes used: R in combination with the RStudio interface.

[1] This paper is an econometrics project, written as part of the Applied Statistics and Econometrics module on the Birkbeck, University of London GDip Economics programme.

1. INTRODUCTION

The study of subjective wellbeing is of interest to economists as it can be used as an empirical approximation for utility. Economists have traditionally measured wellbeing through identifying the resources economic agents have at their command. Given that many of the determinants of human wellbeing are related to life circumstances, and cannot be described by monetary resources or prices, this approach is not adequate. Following recommendations of the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz, Sen & Fittousi, 2009), subjective wellbeing measures have been included in government surveys and are furthering understanding of what contributes to quality of life beyond income or assets or the consumption of goods and services. This project will investigate the bivariate relationship between local authority district mean estimates for happiness and the proportion of their population resident in rural areas in England and Scotland. The local authority estimates are from the Office for National Statistics (ONS) 2011/12 headline estimates for personal wellbeing dataset. The population data are from the Department for Environment, Food and Rural Affairs (DEFRA) 2011 rural-urban classification of local authority districts and the Scottish Government mid-2010 population estimates.

The structure of the project is as follows: 2 explains subjective wellbeing and gives a review of literature; 3 explains the data used in the project; 4 explains the results, in which an association between sample variables is found; and 5 gives a summary of the results.

2. BACKGROUND

Subjective wellbeing is understood to encompass three components: life satisfaction, positive affect, and negative affect. Life satisfaction describes a person's judgment of their life at the moment of measurement. Positive and negative affect constitute the hedonic balance, namely the balance of pleasantness and unpleasantness at any given moment (Diener, 1994). Hedonic measures of subjective-wellbeing reflect a view of wellbeing espoused by Jeremy Bentham, who argued that pleasure and pain are the only things that are good for anyone. According to this view, what enjoyable experiences have in common 'is their positive feeling tone: an intrinsic, unanalysable quality of pleasantness which is present to a greater degree in all of them' (Crisp, 2006). Positive affect can be thought of as happiness or joy, whereas negative affect describes feelings of sadness or anxiety. The psychological literature has not found a relationship between positive and negative affect (Diener, 1984). As such the presence of positive affect in an individual is not an indicator of an absence or a presence of negative affect. The literature does observe a weak correlation (0.40) between life satisfaction and positive affect (Stiglitz et al., 2009).

A survey (Dolan, Peasgood & White, 2008) of economic literature on the factors associated with subjective wellbeing found evidence of rural life being beneficial, and living in large cities being detrimental, to life satisfaction in Sweden (Gerdtham & Johannesson, 2001), Australia (Dockery, 2003), Latin America (Graham & Felton, 2006), and Eastern Europe (Hayo, 2004). Associations between reported levels of life satisfaction and living in rural areas have been observed in the USA (Berry & Okulicz-Kozaryn, 2011), Scotland (Colley, Gilbert & Roberts, 2016), and the UK (ONS, 2018). These analyses, with the exception of USA study, found area effects: how much more or less survey respondents rated their wellbeing compared to a reference group. They also found associations – with happiness (ONS, 2018), and hedonic measures of subjective wellbeing (Colley et al., 2016) – that were not statistically significant. The USA study found that as the environment becomes less dense and more rural, residents reported higher levels of subjective wellbeing. The literature also provides evidence that the variation in subjective wellbeing can be explained by age, gender, employment status, marital status, health, community involvement, safety and deprivation of the area, religion, and personality (Dolan et al., 2008).

3. DATA

Subjective Wellbeing

The Annual Population Survey (APS) includes four questions designed by Dolan, Layard and Metcalfe (2011) to measure life satisfaction, happiness, anxiety, and worthwhile (**app 2.1**). The question measuring happiness is designed to measure the level of positive affect (Dolan et al., 2011).

> *Question 70 - Overall, how happy did you feel yesterday? (Where 0 is 'not happy at all' and 10 is completely happy)* (GSS, 2015).

Of the surveys that measure subjective wellbeing, the APS has been chosen for its inclusion of a geographical component that allows for comparisons between local authority districts and its large sample size, ensuring local authority mean estimates for wellbeing are precise. The sample size for APS Question 70 in 2011/12 in England and Scotland is 119,810 and 23,280 respectively (ONS, 2019).

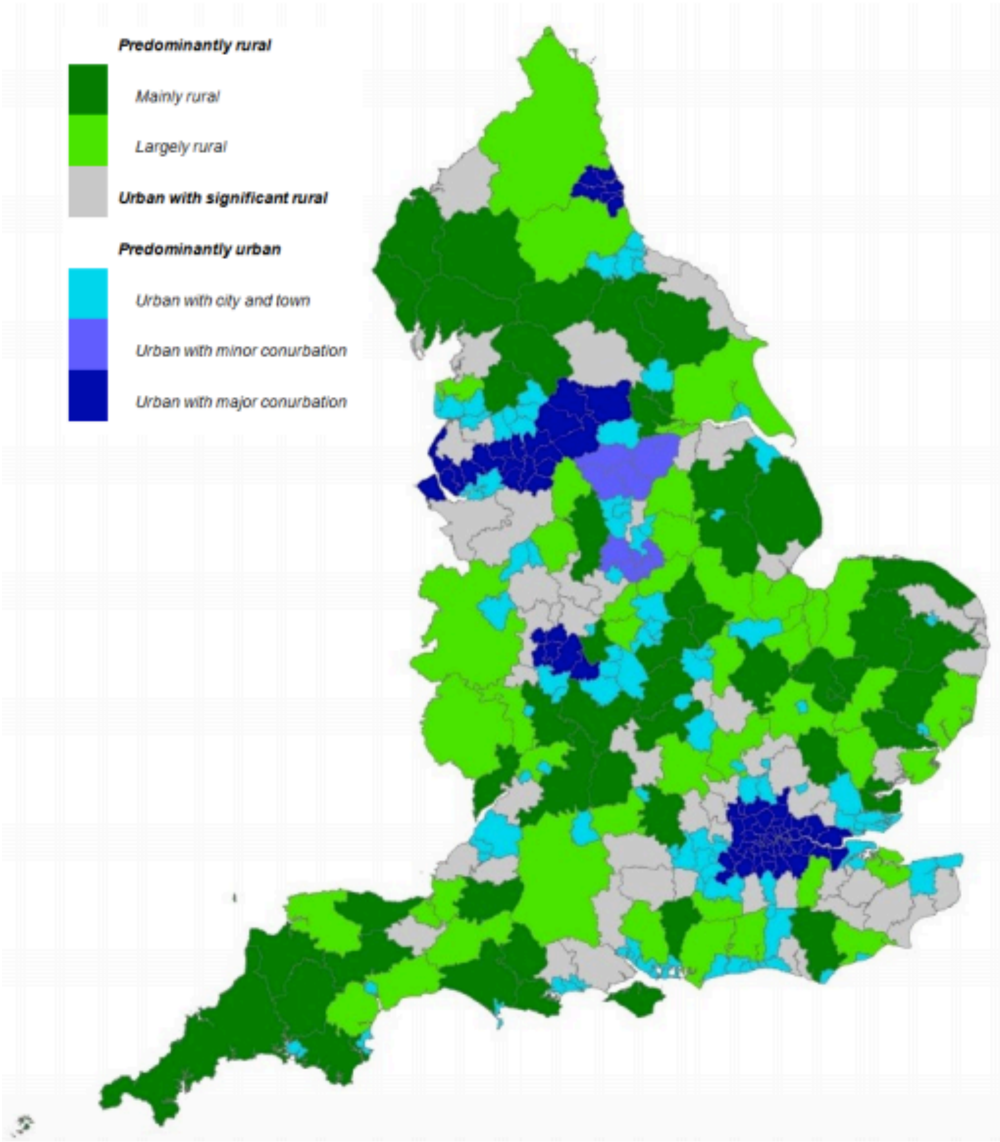
Individual responses to the APS are collected by interview over the telephone or face-to-face in the respondent's home (GSS, 2015). The ONS calculates estimates for mean ratings of happiness for every UK local authority using individual response data from the above question. Given that these data are averages, it should not be inferred that all people in a local authority with a given mean estimate are happier than those in local authority with a lower mean estimate.

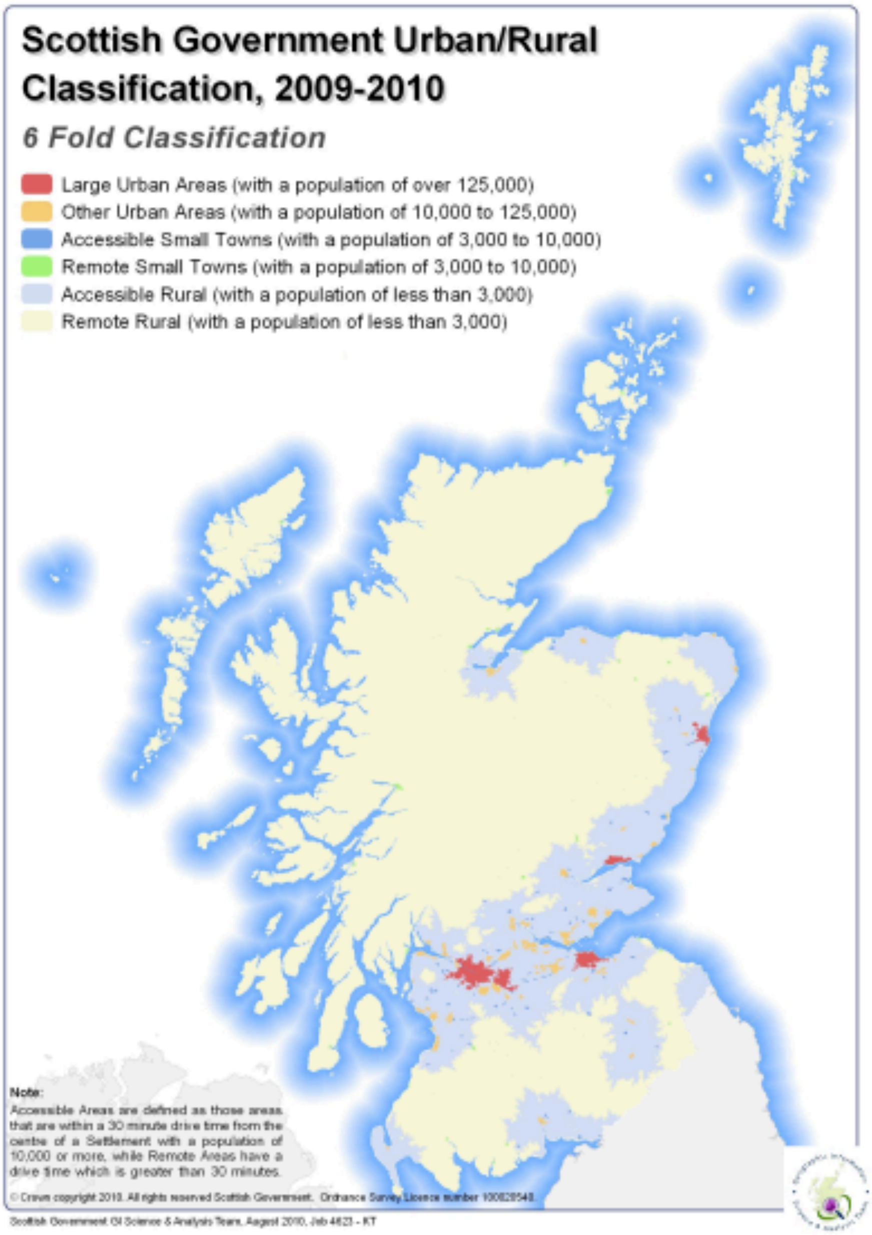
Rural Areas

The Scottish Government defines a rural area as a settlement where fewer than 3000 people live (Scottish Government, 2012). Of the categories in its 6-fold rural-urban classification, the proportion of local authority population under categories 5 and 6 (Accessible Rural and Remote Rural respectively), which are consistent with the Scottish Government's definition of rural areas, will be used in this project (**app 2.2**). These data are part of Scotland's mid-2010 population estimates, which are compiled using postcode population estimates (Scottish Government, 2011).

DEFRA builds the rural-urban classification of local authority districts (LAD) from Census Output Areas. Such areas are defined as rural when their dominant physical settlement has a resident population of fewer than 10,000 people (DEFRA, 2017). The proportion of English LAD population in Output Areas falling under this definition will be used in this project (**app 2.3**). It is important to note that these data do not describe the physical landscape in a given local authority, but the extent to which the resident population live in rural settlements. These data have been compiled using 2011 Census data (DEFRA, 2017a).

Fig.1: DEFRA English Urban-Rural local authority map and Scottish Urban-Rural Classification





Source: DEFRA; Scottish Government

While the Northern Ireland Statistics and Research Agency and the Welsh Government have rural-urban classifications, they do not publish data on proportion of local authority population resident in rural areas. As such, this project excludes LADs in these countries.

Data used in analysis

The combined sample of local authority mean estimates for happiness and proportion of LAD population resident in rural areas total 356 observations. Rural area data are positively skewed (**Table 1**), with over half of local authority districts in England and Scotland having fewer than 18% of their population living in rural areas. Mean estimates for happiness data have an excess kurtosis of 0.49, reflecting a greater proportion of mean estimates in the tails of the distribution than would be expected if the data were normally distributed.

Table 1: Variable descriptive statistics

LAD mean estimates for happiness (out of 10)		**Proportion of LAD population resident in rural areas (%)**	
Mean	7.34	Mean	30.88
Std.Dev	0.23	Std.Dev	33.67
Min	6.74	Min	0.00
Q1	7.18	Q1	0.85
Median	7.31	Median	17.75
Q3	7.47	Q3	52.92
Max	8.07	Max	100.00
MAD	0.21	MAD	26.32
IQR	0.29	IQR	51.95
CV	0.03	CV	1.09
Skewness	0.45	Skewness	0.89
SE.Skewness	0.13	SE.Skewness	0.13
Kurtosis	0.49	Kurtosis	-0.59
N	356	N	356
Source: ONS		Source: DEFRA; Scottish Government	

Across the entire APS sample, respondents interviewed over the telephone scored their happiness on average 0.2 out of 10 higher than those interviewed face-to-face (ONS, 2013). Due to their remoteness, interviews in the Shetland Isles, Western Isles, and the Orkney Islands took place on the telephone only and are thus subject to mode effects (Day and Clements, 2019). All observations, with exception to estimates for the Isles of Scilly, Newcastle-under-Lyme, and the City of London have coefficients of variation below 5%. The mean estimate for Newcastle-under-Lyme has a coefficient of variation between 5% and 10%. The ONS has suppressed estimates for the City of London and the Isles of Scilly, as both have coefficients of variation over 20%, so they will not be accounted for in this project.

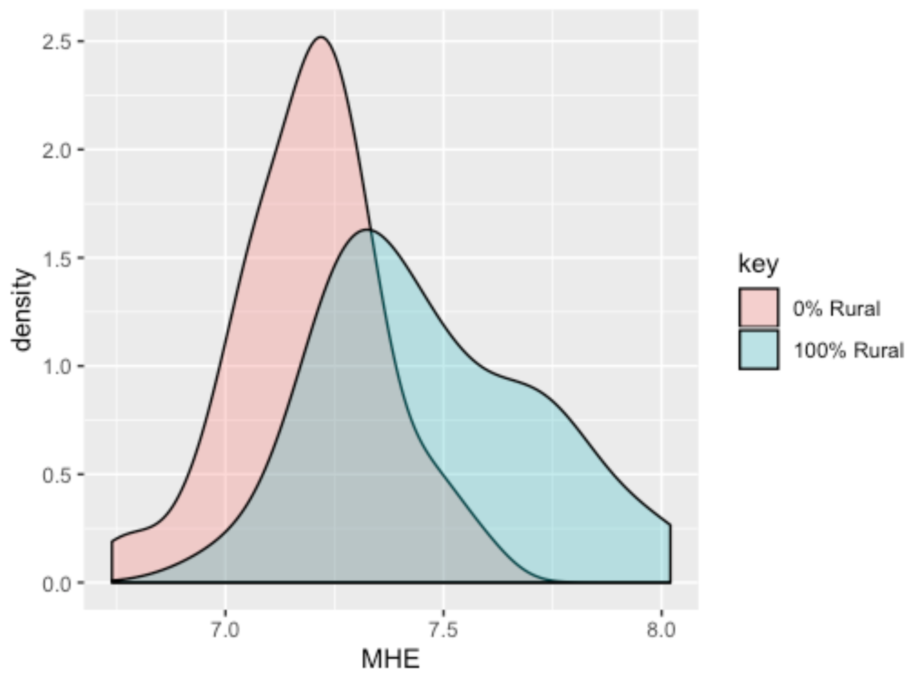
Within the sample, there are 26 LADs that are 100% rural and 52 that are 0% rural. Mean estimates at the 25^{th}, 50^{th}, and 75^{th} percentile are higher in 100% rural than in 0% rural subsample (**Table 2**). A density plot of the two subsamples (**fig.2**) show that the mean estimates for 100% rural LADs are positively skewed, and are more dispersed than the mean estimates for 0% rural LADs that are negatively skewed. Across the entire sample used in this project, the LAD with the lowest mean estimate for happiness (6.74), Wolverhampton, falls within the 0% rural subsample. The LAD with the highest mean estimate (8.07), Chichester, falls outside of the 100% subsample with 64% of is population resident in rural areas.

Table 2: Descriptive statistics for LAD mean estimates for happiness for 0% and 100% Rural LADs

0% Rural (out of 10)		100% Rural (out of 10)	
Mean	7.19	Mean	7.47
Std.Dev	0.17	Std.Dev	0.25
Min	6.74	Min	7.01
Q1	7.08	Q1	7.29
Median	7.21	Median	7.43
Q3	7.29	Q3	7.68
Max	7.60	Max	8.02
MAD	0.15	MAD	0.24
IQR	0.21	IQR	0.39
CV	0.02	CV	0.03
Skewness	-0.16	Skewness	0.44
SE.Skewness	0.33	SE.Skewness	0.46
Kurtosis	0.30	Kurtosis	-0.70
N	52	N	26

Source: ONS

Fig.2: Density plots of LAD mean estimates for happiness (out of 10) of 0% and 100% Rural LADs



Source: ONS

4. ANALYSIS

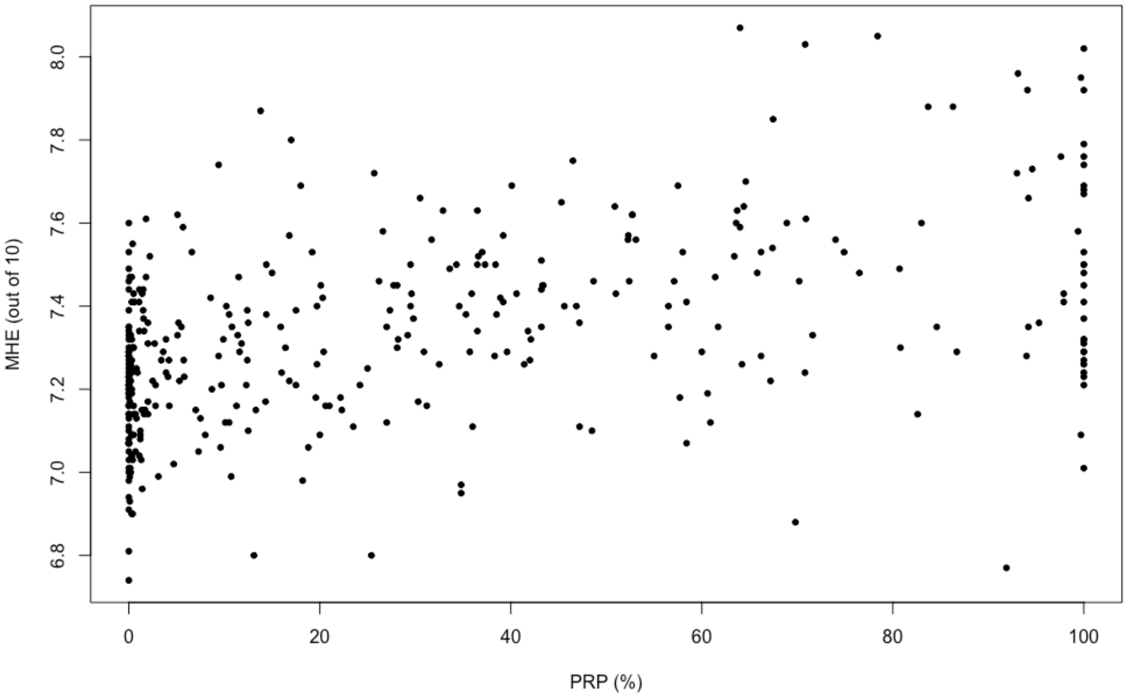
Model specification

Data were compiled using Excel and analyses were carried out using R in combination with the RStudio interface.

A plot of LAD mean estimates for happiness (MHE) and the proportion of the LAD population resident in rural areas (PRP) suggest an association between the two (**fig.3**). To investigate the linear association, a bivariate model was specified:

$$MHE_i = \alpha + \beta PRP_i + u_i$$

Fig.3: Plot of MHE vs PRP



Source: ONS, DEFRA, Scottish Government

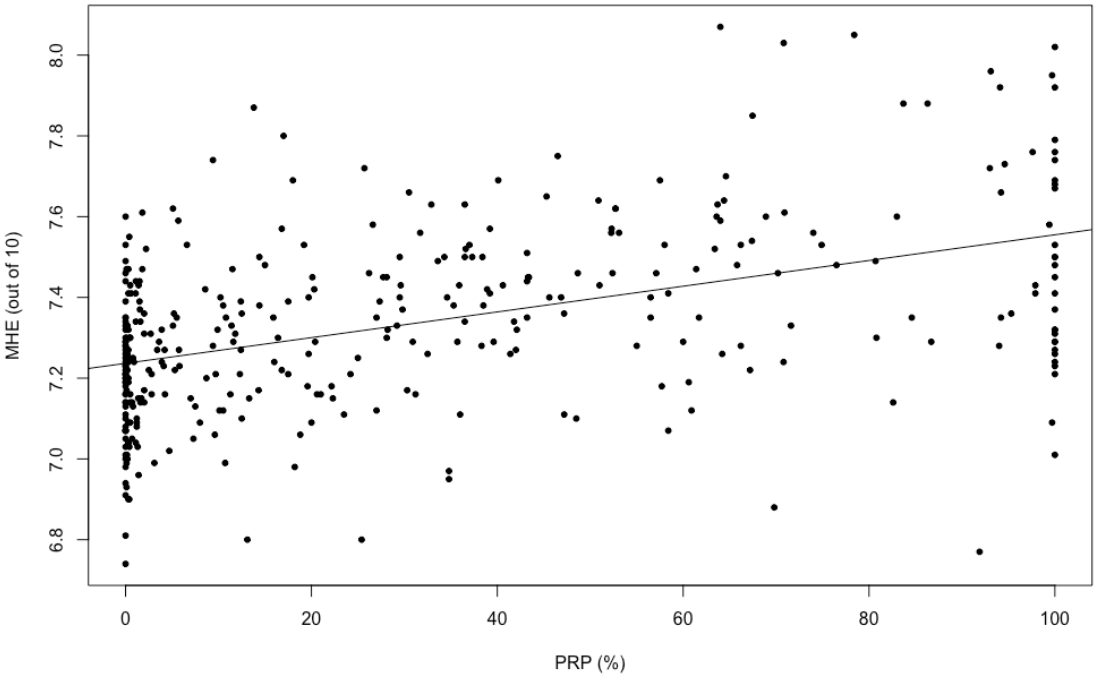
Estimates for α and β are as follows (**app 1.1**) (**fig.4**):

$$MHE_i = 7.236 + 0.00318 PRP_i + \hat{u}_i$$
$$(0.0131) \quad (0.000383)$$

$N = 356 \quad R^2 = 0.2155 \quad SER = 0.2049 \quad df = 354$

	t-ratio	p value	lower	upper
(Intercept)	550.3522	< 2.2e-16	7.21094	7.26249
PRP	8.3105	2.063e-15	0.00243421	0.00393674

Fig.4: Plot of MHE vs PRP with OLS regression line



Source: ONS, DEFRA, Scottish Government

Tests carried out on the above model provide evidence of heteroscedasticity (**app 1.2**) and misspecification (**app 1.3**). As a result White standard errors have been used (**app 1.4**). Heteroscedasticity may have been caused by omission of explanatory variables. It could also indicate that the association between MHE and PRP is not constant across the data set. To fix heteroscedasticity and misspecification issues, subsets of the data for values 0<PRP<50 and 50<PRP<100 were created, and MHE values were regressed by PRP values in each dataset:

0<PRP<50: $MHE_i = \alpha + \beta \underset{0-50}{PRP}_i + u_i$

50<PRP<100: $MHE_i = \alpha + \beta \underset{50-100}{PRP}_i + u_i$

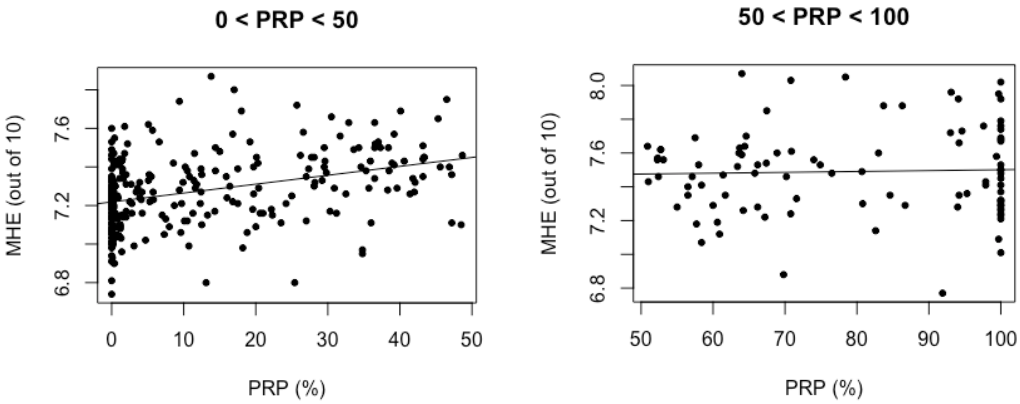
The samples for 0<PRP<50 and 50<PRP<100 total 260 and 96 observations respectively. There is a statistically significant association between MHE and PRP for LADs that have less than 50% of their population resident in rural areas (**app 1.5**). The association between MHE and PRP for LADs with more than 50% of their population in rural areas is not statistically significant, which suggests that variations in PRP above the 50% threshold have no influence in predicting the average level of happiness (**Table 3**)(**app 1.6**). MM estimates computed for the 0<PRP<50 model were 7.214 for the intercept and 0.004799 for PRP coefficient, indicating that OLS estimates have not been biased by outliers (**app 1.7**). Robust estimates were not calculated for the 50<PRP<100 model as there is no statistically significant association between MHE and PRP.

Table 3: OLS estimates of α and β coefficients

	0<PRP<50			50<PRP<100		
	Estimate	Std. Error	t-ratio	Estimate	Std. Error	t-ratio
(Intercept)	7.21839	0.0146194	493.754*	7.44979	0.120076	62.043*
PRP	0.0046134	0.0007412	6.225*	0.000511	0.00147	0.348
	N=260			N=96		
	R^2=0.1306	SER: 0.1788	df=258	R^2=0.001284	SER=0.2583	df=94

* Coefficient is individually significant at the 1% level

Fig.5: Plots of MHE vs PRP with OLS regression line

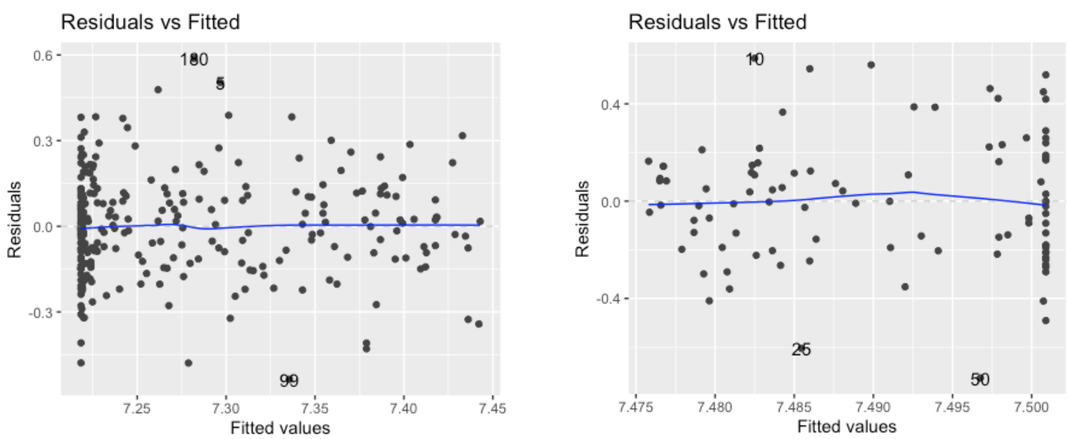


Source: ONS; DEFRA; Scottish Government

Plots of residuals and fitted values for both models (**fig.6**) show the residuals scattered around the zero line, indicating that for the 0<PRP<50 sample the assumption of a linear relationship is reasonable. Outliers in the 0<PRP<50 residual plot are observations 99 (Guildford) and 180 (Rochford) whose MHE values are poorly predicted by the model. Tests carried out on both models provide evidence of normally distributed and homoscedastic residuals, and no misspecification (**app 1.8 – 1.13**).

Fig.6: Residual plots

0<PRP<50 **50<PRP<100**



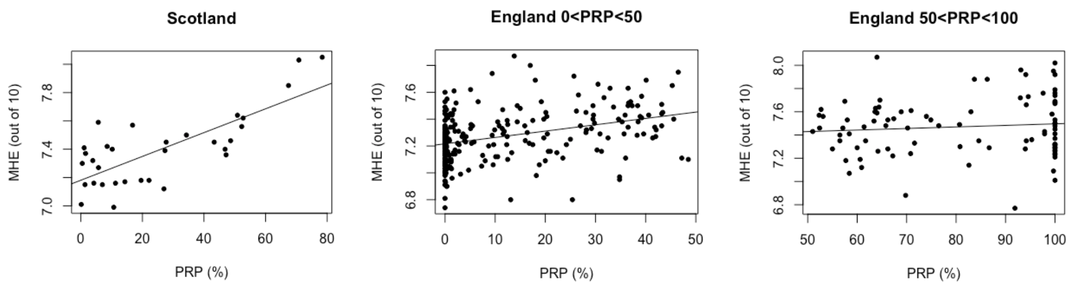
Comparisons between England and Scotland

The above models assume that, to the extent that a statistically significant association between MHE and PRP has been found, the strength of the association is constant across England and Scotland. To test this MHE was regressed by PRP for all LADs in Scotland and in English LADs such that 0<PRP<50 and 50<PRP<100:

Scotland: $MHE_i = \alpha^S + \beta^S PRP_i + u_i$;

England: $MHE_i = \alpha^E + \beta^E \underset{0-50}{PRP}{}_i + u_i$; $MHE_i = \alpha^E + \beta^E \underset{50-100}{PRP}{}_i + u_i$.

Fig.7: Plots of MHE vs PRP with OLS regression line



Source: ONS; Scottish Government Source: ONS; DEFRA

The samples for Scotland, England 0<PRP<50 and England 50<PRP<100 total 32, 234, and 90 observations respectively. The results (**Table 4**)(**app 1.14 - 1.16**) indicate that the assumption doesn't hold. Of the statistically significant associations found, the association is stronger in Scotland than in England 0<PRP<50. For England 50<PRP<100, there is no statistically significant association between MHE and PRP. The variation in the strength of the association between the two countries suggests that the proportion of LAD population in rural areas has greater influence on average level of happiness in Scotland than it does in England.

Table 4: OLS estimates of α and β coefficients

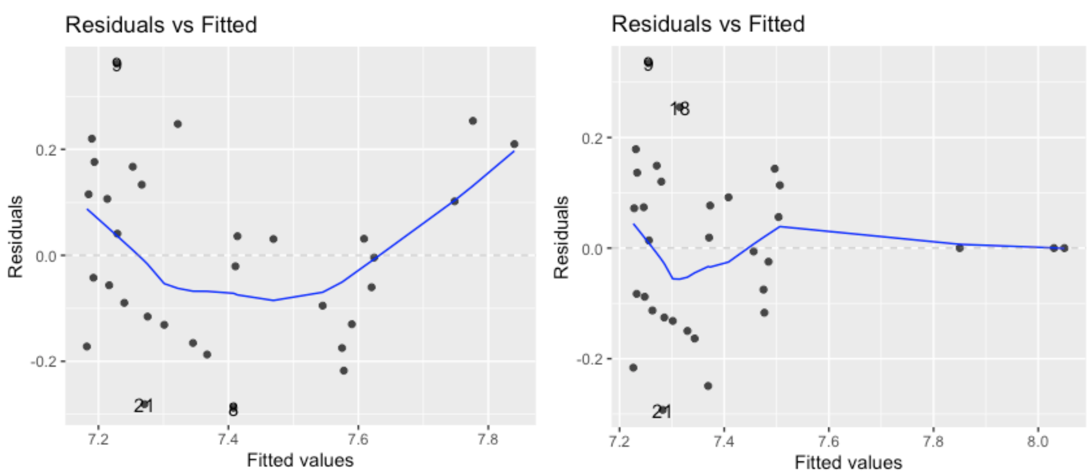
	Scotland			**Scotland with dummies**		
	Estimate	Std. Error	t-ratio	Estimate	Std. Error	t-ratio
(Intercept)	7.18085	0.045151	159.040*	7.22543	0.043781	165.036*
PRP	0.008408	0.001313	6.404*	0.005327	0.001585	3.361*
Western Isles				0.40690	0.18234	2.232
Orkney Islands				0.26520	0.17430	1.522
Shetland Isles				0.42734	0.17662	2.419
	N=32			N=32		
	R^2=0.5775	SER=0.1704	df=30	R^2=0.6846	SER=0.1552	df=27
England	**0<PRP<50**			**50<PRP<100**		
	Estimate	Std. Error	t-ratio	Estimate	Std. Error	t-ratio
(Intercept)	7.21628	0.0153665	466.576*	7.36294	0.12227	60.22*
PRP	0.0047190	0.0008015	5.888*	0.001329	0.001477	0.9
	N=234			N=90		
	R^2=0.13	SER=0.1816	df=232	R^2=0.009121	SER=0.2481	df=88

* Coefficient is individually significant at the 1% level

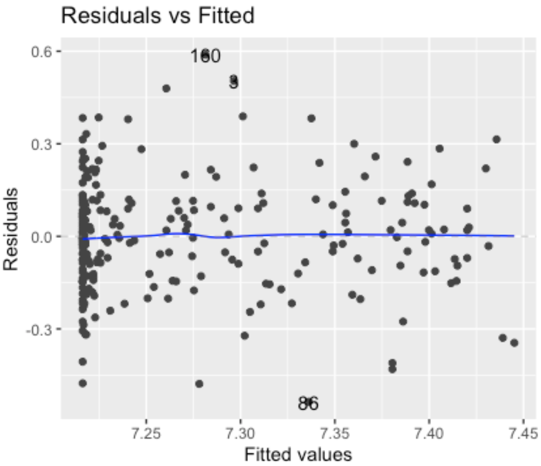
Residual plots for the England models show the data points scattered randomly around the zero line (**fig. 8**). Tests carried out on the three models specified above provide evidence of normally distributed, and homoscedastic residuals (**app. 1.17-1.22**). In the residual plot of the Scotland model the data points have a discernable pattern. This pattern can be explained in part due to the three observations with the highest fitted values corresponding to LADs – Western Isles, Shetland Isles, and the Orkney Islands – whose mean estimates for happiness are subject to mode effects. The result of regressing MHE by PRP with dummy variables for these three observations is a PRP coefficient of 0.005327 (**app. 1.23**), suggesting that the three observations have distorted the extent of the difference in the strength of the association between the Scotland and England 0<PRP<50 models.

Fig.8: Residual plots

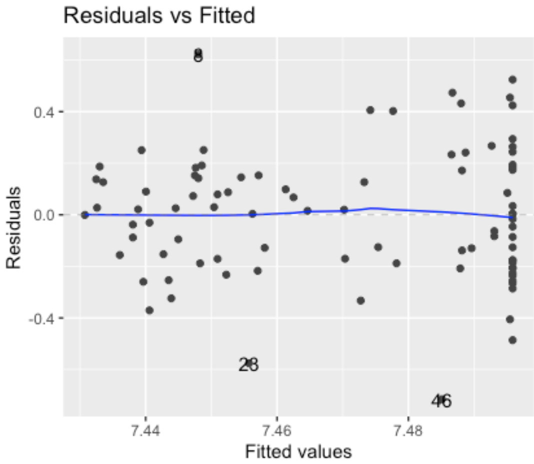
Scotland **Scotland with dummies**



England 0<PRP<50



England 50<PRP<100



5. SUMMARY & CONCLUSIONS

Across England and Scotland, there is a statistically significant association between local authority mean estimates for happiness and the share of local authority population in rural areas for local authority districts that are less than 50% rural. The analysis found no statistically significant association in local authorities that are more than 50% rural. The result accords in part with Berry and Okulicz-Kozaryn's findings, but differs in that the association in the USA was observed across all areas from dense urban areas to remote rural areas.

Analyses of subsets of the sample relating to Scottish and English local authorities found statistically significant associations of varying strengths. The association between sample variables is stronger across Scottish local authorities compared to English local authorities that are less than 50% rural. There is no statistically significant association between sample variables in English local authorities that are more than 50% rural. The result for Scotland is in contrast to the Colley, Gilbert and Roberts analysis that did not find a statistically significant association between hedonic wellbeing and living in rural areas.

It is important to note that where a statistically significant association has been found, this result should not be taken to imply a causal link between sample variables. It cannot be inferred from the results that an individual within a given local authority will be happier than another in a local authority with a lower proportion of its population in rural areas. Similarly, the results do not show that all people living within a more rural local authority will be happier than those in a less rural local authority nor do they predict that moving to a more rural local authority will make a given individual happier. The results do predict, however, that on average people will be happier in a more rural local authority.

The bivariate models used in this project are accurate in predicting mean estimates for happiness in 2011/12 in English local authorities that are less than 50% rural and Scottish local authorities to within 0.18 and 0.17 out of 10 respectively. Further work could improve this through regressing local authority mean estimates for happiness with other variables known to have an association with subjective wellbeing such as unemployment rates, crime rates, and pollution. This would further the understanding of the source of the benefit of living in rural areas and contribute to the development of policies that aim to recreate that benefit in urban areas.

REFERENCES

Berry, Brian J. L., and Adam Okulicz-Kozaryn. "An Urban-Rural Happiness Gradient." *Urban Geography* (Routledge) 36, no. 6 (2011): 871-883.

Colley, Kathryn, Alana Gilbert, and Deborah Roberts. "Are rural residents happier? A quantitative analysis of wellbeing in Scotland." *Journal of Rural Studies* 44 (April 2016): 37-45.

Crisp, Roger. "Hedonism Reconsidered." *Philosophy and Phenomenological Research* LXXIII, no. 3 (2006): 619-645.

Day, Laurence, and Liam Clements. *Personal well-being in the UK: April 2018 to March 2019.* ONS, 2019.

DEFRA. *2011 Rural Urban Classification of Local Authority Districts and Similar Geographic Units in England: A User Guide.* ONS, 2017.

DEFRA. *2011 Rural-Urban Classification of Local Authority Districts and Similar Geographic Units in England: Methodology.* ONS, 2017a.

Diener, Ed. "Assesing Subjective Well-Being: Progress and Opportunities." *Social Indicators Research* (Springer) 31, no. 2 (February 1994): 103-157.

Diener, Ed. "Subjective Well-Being." *Psychological Bulletin* (American Psychological Association) 95, no. 3 (1984): 542-553.

Dockery, Alfred Michael. "Happiness, life satisfaction and the role of work: Evidence from two Austrailian surveys." 2003.

Dolan, Paul, Richard Layard, and Robert Metcalfe. *Measuring Subjective Well-being for Public Policy Purposes.* ONS, 2011.

Dolan, Paul, Tessa Peasgood, and Mathew White. "Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being." *Journal of Economic Psychology* 29, no. 1 (February 2008): 94-122.

Gerdtham, Ulf-G, and Magnus Johannesson. "The relationship between happiness, health, and socio-economic factors: results based on Swedish microdata." *The Journal of Socio-Economics* (Elsevier) 30, no. 6 (November-December 2001): 553-557.

Graham, Carol, and Andrew Felton. "Inequality and happiness: Insights from Latin America." *Journal of Economic Inequality* 4 (2006): 107-122.

GSS. *Harmonised Concepts and Questions for Social Data Sources: Personal Well-being.* Newport: Government Statistical Service, 2015.

Hayo, Bernd. "Happiess in Eastern Europe." (Philipps-University Marburg and ZEI) 2004.

ONS. "APS Regression Models October 2017 to September 2018." ONS, 2018.

—. "Personal Well-being in the UK 2012/13." 2013.

—. "Quality information for personal wellbeing October 2018 to September 2019." ONS, 2019.

Scottish Government. *Percentage population estimate by Urban Rural Classification.* Edinburgh, 2011.
Scottish Government. *Scottish Government Urban/Rural Classification.* Scottish Government, 2012.
Stiglitz, Joseph E, Armatya Sen, and Jean-Paul Fitoussi. "Report by the Commission on the Measurement of Economic Performance and Social Progress." 2009, 143-232.

APPENDICES

1. Outputs from R

1.1: OLS regression output

```
Call:
lm(formula = MHE ~ PRP, data = ASE_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.75947 -0.13803  0.00034  0.12479  0.62941 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.236720   0.014745 490.784   <2e-16 ***
PRP         0.003185   0.000323   9.862   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2049 on 354 degrees of freedom
Multiple R-squared:  0.2155,	Adjusted R-squared:  0.2133 
F-statistic: 97.25 on 1 and 354 DF,  p-value: < 2.2e-16
```

1.2: F Test for heteroscedasticity

```
 F Test for Heteroskedasticity
 -----------------------------
 Ho: Variance is homogenous
 Ha: Variance is not homogenous

 Variables: fitted values of MHE

       Test Summary
 ----------------------------
 Num DF     =    1
 Den DF     =    354
 F          =    22.0315
 Prob > F   =    3.840028e-06
```

1.3: Model reset test

```
	RESET test

data:  model
RESET = 5.2962, df1 = 2, df2 = 28, p-value = 0.0112
```

1.4: White standard errors

```
t test of coefficients:

              Estimate Std. Error  t value  Pr(>|t|)    
(Intercept) 7.23672049 0.01314925 550.3522 < 2.2e-16 ***
PRP         0.00318548 0.00038331   8.3105 2.063e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

1.5: 0<PRP<50 model regression output

```
Call:
lm(formula = ZeroToFiftyPRP$MHE ~ ZeroToFiftyPRP$PRP, data = ZeroToFiftyPRP)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.53558 -0.11592  0.00137  0.10834  0.58794 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)        7.2183959  0.0146194 493.754  < 2e-16 ***
ZeroToFiftyPRP$PRP 0.0046134  0.0007412   6.225 1.94e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1788 on 258 degrees of freedom
Multiple R-squared:  0.1306,	Adjusted R-squared:  0.1272 
F-statistic: 38.75 on 1 and 258 DF,  p-value: 1.944e-09
```

1.6: 50<PRP<100 model regression output

```
Call:
lm(formula = FiftyToHundredPRP$MHE ~ FiftyToHundredPRP$PRP, data = FiftyToHundredPRP)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.72676 -0.19094 -0.00615  0.15849  0.58750 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           7.449792   0.120076  62.043   <2e-16 ***
FiftyToHundredPRP$PRP 0.000511   0.001470   0.348    0.729    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2583 on 94 degrees of freedom
Multiple R-squared:  0.001284,	Adjusted R-squared:  -0.00934 
F-statistic: 0.1209 on 1 and 94 DF,  p-value: 0.7289
```

1.7: MM estimates for 0<PRP<50 model

```
Call:
lmrob(formula = ZeroToFiftyPRP$MHE ~ ZeroToFiftyPRP$PRP)
 \--> method = "MM"
Residuals:
      Min        1Q    Median        3Q       Max 
-0.536656 -0.115736  0.004987  0.108071  0.589007 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)        7.2147733  0.0140938 511.911  < 2e-16 ***
ZeroToFiftyPRP$PRP 0.0047986  0.0007085   6.773  8.5e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Robust residual standard error: 0.1673 
Multiple R-squared:  0.1508,	Adjusted R-squared:  0.1475 
Convergence in 6 IRWLS iterations
```

1.8: 0<PRP<50 model RESET test

```
        RESET test

data:  ZeroToFiftyPRP$MHE ~ ZeroToFiftyPRP$PRP
RESET = 0.55295, df1 = 2, df2 = 256, p-value = 0.5759
```

1.9: 50<PRP<100 model RESET test

```
        RESET test

data:  FiftyToHundredPRP$MHE ~ FiftyToHundredPRP$PRP
RESET = 0.58423, df1 = 2, df2 = 92, p-value = 0.5596
```

1.10: F test for heteroscedasticity for 0<PRP<50 model

```
F Test for Heteroskedasticity
-----------------------------
Ho: Variance is homogenous
Ha: Variance is not homogenous

Variables: fitted values of ZeroToFiftyPRP$MHE

     Test Summary
-------------------------
Num DF     =    1
Den DF     =    258
F          =    0.2111002
Prob > F   =    0.6462933
```

1.11: F test for heteroscedasticity for 50<PRP<100 model

```
F Test for Heteroskedasticity
-----------------------------
Ho: Variance is homogenous
Ha: Variance is not homogenous

Variables: fitted values of FiftyToHundredPRP$MHE

     Test Summary
-------------------------
Num DF     =    1
Den DF     =    94
F          =    1.955079
Prob > F   =    0.1653309
```

1.12: Jarque-Bera test for 0<PRP<50 model residuals

```
        Jarque-Bera Normality Test

data:  residuals(ZeroToFiftyModel)
JB = 3.1337, p-value = 0.2087
alternative hypothesis: greater
```

1.13: Jarque-Bera test for 50<PRP<100 model residuals

```
        Jarque-Bera Normality Test

data:  residuals(FiftyToHundredModel)
JB = 0.082404, p-value = 0.9596
alternative hypothesis: greater
```

1.14: Scotland regression output

```
Call:
lm(formula = ASE_data_Scotland$MHE ~ ASE_data_Scotland$PRP, data = ASE_data_Scotland)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.28790 -0.13030 -0.01251  0.11968  0.36146 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           7.180853   0.045151 159.040  < 2e-16 ***
ASE_data_Scotland$PRP 0.008408   0.001313   6.404 4.54e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1704 on 30 degrees of freedom
Multiple R-squared:  0.5775,	Adjusted R-squared:  0.5634 
F-statistic: 41.01 on 1 and 30 DF,  p-value: 4.536e-07
```

1.15: England 0<PRP<50 regression output

```
Call:
lm(formula = EnglandZeroToFifty$MHE ~ EnglandZeroToFifty$PRP, 
    data = EnglandZeroToFifty)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.53615 -0.11962  0.00301  0.10775  0.58859 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)            7.2162843  0.0154665 466.576  < 2e-16 ***
EnglandZeroToFifty$PRP 0.0047190  0.0008015   5.888 1.36e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1816 on 232 degrees of freedom
Multiple R-squared:   0.13,	Adjusted R-squared:  0.1262 
F-statistic: 34.67 on 1 and 232 DF,  p-value: 1.363e-08
```

1.16: England 50<PRP<100 regression output

```
Call:
lm(formula = EnglandFiftyToHundred$MHE ~ EnglandFiftyToHundred$PRP, 
    data = EnglandFiftyToHundred)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.71510 -0.17587  0.00413  0.16708  0.62198 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)               7.362948   0.122270   60.22   <2e-16 ***
EnglandFiftyToHundred$PRP 0.001329   0.001477    0.90    0.371    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2481 on 88 degrees of freedom
Multiple R-squared:  0.009121,	Adjusted R-squared:  -0.002139 
F-statistic:  0.81 on 1 and 88 DF,  p-value: 0.3706
```

1.17: Scotland model Jarque-Bera test

```
        Jarque-Bera Normality Test

data:  residuals(scotland)
JB = 1.1044, p-value = 0.5757
alternative hypothesis: greater
```

1.18: England 0<PRP<50 model Jarque-Bera test

```
        Jarque-Bera Normality Test

data:  residuals(EnglandZeroFiftyModel)
JB = 3.12, p-value = 0.2101
alternative hypothesis: greater
```

1.19: England 50<PRP<100 model Jarque-Bera test

```
        Jarque-Bera Normality Test

data:  residuals(EnglandFiftyHundredModel)
JB = 0.10301, p-value = 0.9498
alternative hypothesis: greater
```

1.20: F test for heteroscedasticity for Scotland model

```
F Test for Heteroskedasticity
-----------------------------
Ho: Variance is homogenous
Ha: Variance is not homogenous

Variables: fitted values of ASE_data_Scotland$MHE

     Test Summary
-------------------------
Num DF     =    1
Den DF     =    30
F          =    0.08749411
Prob > F   =    0.7694252
```

1.21: F test for heteroscedasticity for England 0<PRP<50 model

```
F Test for Heteroskedasticity
-----------------------------
Ho: Variance is homogenous
Ha: Variance is not homogenous

Variables: fitted values of EnglandZeroToFifty$MHE

     Test Summary
-------------------------
Num DF     =    1
Den DF     =    232
F          =    0.7126854
Prob > F   =    0.3994234
```

1.22: F test for heteroscedasticity for England 50<PRP<100

```
F Test for Heteroskedasticity
-----------------------------
Ho: Variance is homogenous
Ha: Variance is not homogenous

Variables: fitted values of EnglandFiftyToHundred$MHE

     Test Summary
-------------------------
Num DF     =    1
Den DF     =    88
F          =    2.599867
Prob > F   =    0.1104535
```

1.23: Scotland with dummy variables regression output

```
Call:
lm(formula = ASE_Scotland_dummies$MHE ~ ASE_Scotland_dummies$PRP +
    ASE_Scotland_dummies$`LAD_Na h-Eileanan Siar` + ASE_Scotland_dummies$`LAD_Orkney Islands` +
    ASE_Scotland_dummies$`LAD_Shetland Islands`)

Residuals:
    Min      1Q  Median      3Q     Max
-0.2925 -0.1138  0.0000  0.0973  0.3343

Coefficients:
                                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                                   7.225439   0.043781 165.036  < 2e-16 ***
ASE_Scotland_dummies$PRP                      0.005327   0.001585   3.361  0.00233 **
ASE_Scotland_dummies$`LAD_Na h-Eileanan Siar` 0.406909   0.182344   2.232  0.03414 *
ASE_Scotland_dummies$`LAD_Orkney Islands`     0.265205   0.174304   1.522  0.13976
ASE_Scotland_dummies$`LAD_Shetland Islands`   0.427340   0.176626   2.419  0.02255 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1552 on 27 degrees of freedom
Multiple R-squared:  0.6846,	Adjusted R-squared:  0.6379
F-statistic: 14.65 on 4 and 27 DF,  p-value: 1.756e-06
```

2. Variable Definitions

2.1: Subjective Wellbeing Estimates

ONS Personal wellbeing measures ask people to evaluate, on a scale of 0 to 10, how satisfied they are with their live overall, whether they feel they have meaning and purpose in their life, and about their emotions (happiness and anxiety) during a particular period.

The four personal wellbeing questions are:

- Overall, how satisfied are you with your life nowadays?
- Overall, to what extent do you feel the things you do in your life are worthwhile?
- Overall, how happy did you feel yesterday?
- Overall, how anxious did you feel yesterday?

Thresholds are used to present dispersion in the data. For life satisfaction, worthwhile and happiness questions, rating are grouped in the following way:

- 0 to 4 (low)
- 5 to 6 (medium)
- 7 to 8 (high)
- 9 to 10 (very high)

Estimates of the mean ratings of all four wellbeing questions are produced at a national level, country and local authority level. See Day and Clements (2019) for extended discussion.

2.2: Scottish 6-fold Rural Urban Classification

1. Large Urban Areas	Settlements of 125,000 or more people.
2. Other Urban Areas	Settlements of 10,000 to 124,999 people.
3. Accessible Small Towns	Settlements of 3,000 to 9,999 people and within 30 minutes drive of a settlement of 10,000 or more.
4. Remote Small Towns	Settlements of 3,000 to 9,999 people and with a drive time of over 30 minutes to a settlement of 10,000 or more.
5. Accessible Rural	Areas with a population of less than 3,000 people, and within a 30 minute drive time of a settlement of 10,000 or more.
6. Remote Rural	Areas with a population of less than 3,000 people, and with a drive time of over 30 minutes to a settlement of 10,000 or more.

2.3: 2011 Rural Urban Classification of Local Authority Districts in England

The urban domain comprises all physical settlements with a population of 10,000 or more. The classification of small areas focuses on the Output Area level - the smallest units for which data are made available from the decennial Population Census. If the majority of the population of a particular small area live in such a settlement, that area is deemed 'urban'; all other Output Areas are deemed 'rural'. See DEFRA (2017) for extended discussion.

For each Local Authority District the Rural Urban Classification provides:

- The name of the authority;
- The number of persons in rural Output Areas (referred to as the 'rural population');
- The number of persons resident in Output Areas which formed part of a Hub Town (referred to as the 'rural related population');
- The sum of the rural and rural related population components;
- The total population usually resident in the district; and
- The combined rural and rural related population components as a percentage of the total population usually resident in the district.

3. Data Sources

The combined sample of local authority mean estimates for happiness and proportion of local authority population resident in rural areas is drawn from:

- **ONS**, Annual personal wellbeing estimates, 'April 2019 to March 2019 – Local authority update' dataset, Happiness by countries, local and unitary authorities tables, 2011/12 Average (mean) ratings column, https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/datasets/headlineestimatesofpersonalwellbeing
- **DEFRA**, Local Authority Districts ranked by rural and hub town (rural related) populations 2011, Percentage of the total population 2011 column, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/591466/Local_Authority_Districts_ranked_by_rural_and_rural-related_populations_with_Rural_Urban_Classification.pdf
- **Scottish Government**, Urban Rural Classification 2011-2012 Population Tables dataset, CA6FOLD tables, 'Accessible rural' and 'Remote rural' columns, https://www2.gov.scot/Topics/Statistics/About/Methodology/UrbanRuralClassification/Urban-Rural-Classification-2011-12/Urban-Rural-2011-2012