KEY FACTORS INFLUENCING THE UK'S REFERENDUM

New Findings on Key Factors Influencing the UK's Referendum on Leaving the EU

Aihua Zhang

Department of Mathematics

University of Leicester

Author Note

Correspondence concerning this article should be addressed to Aihua Zhang,
Department of Mathematics, University of Leicester, University Road, Leicester, LE1 7RH,
United Kingdom

Contact: az90@le.ac.uk

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Abstract

The UK's EU in/out referendum raised significant debate and speculation of the intention of the electorate and its motivations in voting; much of this debate was informed by simple data analysis examining individual factors, in isolation, and using opinion polling data. This, in the case of the EU referendum where multiple factors influence the decision simultaneously, failed to predict the eventual outcome. On June 23, 2016, Britain's vote to leave the EU came as a surprise to most observers, with a bigger voter turnout than that of any UK general election in the past decade. In this research, we apply Multivariate Regression Analysis and a Logit Model to real voting data to identify statistically significant factors influencing the EU referendum voting preference simultaneously as well as the odd ratio in favour of Leave. Visualizations of the key findings are also provided with heat maps and graphs.

We find that higher education is the predominant factor dividing the nation, with a marginal effect on the referendum decision being stronger than any other factors particularly in England and Wales, where most Leave voters reside. An increase of about 3% in the proportion of British adults accessing to higher education in England and Wales could have reversed the referendum result in the UK. We also find that areas in England and Wales with a lower unemployment rate tend to have a higher turnout to support Leave while areas in Scotland and Northern Ireland with a higher proportion of university-educated British adults have a higher turnout to support Remain. Further we find that areas with high proportions of British male adults show a higher percentage of Leave votes. A higher proportion of elderly British contributes to a higher percentage of Leave votes, but does not lead to Leave outcomes on their own.

Keywords: EU Referendum; Turnout; Higher Education; Multivariate Regression; Logit Model; Odds Ratio

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Highlights

- Higher education is found to be the predominant factor dividing the nation between Remain and Leave.
- An increase of about 3% of British adults accessing to higher education in England and Wales could have reversed the referendum result.
- A decrease of about 7% in turnout in England and Wales could have changed the result of the referendum.
- The factor of elderly voters, although having an effect on the outcome, was generally over reported as a dominant factor.
- Sex is found to be a statistically significant factor while British born proportions and the local income levels are insignificant factors.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

1. Introduction

The UK's referendum on whether to remain as a member state of the European Union (EU) caused much debate and speculation at the time of the plebiscite and even more since the results in favour of Leave. High on this list of conjectures was the motivations of the electorate and the voting patterns that this would produce.

There had been significant speculation on the characteristics of people who would either vote to Leave or Remain in the referendum of June 23, 2016 (see, for example, the media articles in *The Economist* (2016) and *The Telegraph* (Kirk, 2016)). Characteristically, it was concluded that the more affluent and better-educated voter would support Remain; whilst the less affluent and those with lower educational opportunities would support Leave; young people were more likely than the old to vote for Remain; old people were more likely to turn out to vote than the young, which was taken to favour Leave. Much of the conjecture, up to and during the referendum, was informed by simple data analysis examining individual factors, in isolation, and using opinion polling data (such as YouGov polling data) and sampling from social media (such as BBC and *The Guardian*). Moreover, there has been an absence of formal academic assessment along this line, although Stolz, Harrington and Porter (2016) have recently conducted research on Brexit but this principally examines the topological 'shape' of it.

The referendum data is now available (thanks to The Electoral Commission's work) and it is therefore, possible to use the actual data to conduct a systematic analysis and provide a more definitive assessment of the key factors that have influenced the British people's decision to leave the EU. However, the referendum data does not contain the voters' personal information, such as sex, education, social grade, income etc., which is required to provide a

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

complete assessment; consequently, we need to use 2011 census¹ data to calculate the corresponding percentages.

Although some studies, in the media, of the voting results have been attempted following the EU referendum (see, for example, BBC News, 2016, The Data Team, 2016a and The Guardian, 2016), the conclusions drawn on the voting patterns appear, mainly, to be based on graphically comparisons of the voting or polling data against each single individual demographic factor. For example, it was concluded, by the BBC News (2016), that older voters were much more likely to vote Leave than younger voters by simply comparing, graphically, the percentage of each age group against the corresponding percentage of Leave/Remain votes using the Lord Ashcroft polling data. The Guardian (2016) compared, graphically, the referendum results to some key demographic characteristics (such as % residents with higher education and % residents not born in the UK) of the local authority areas, individually, to draw some patterns without mentioning what exact data were being used. These comparisons for each individual factor in isolation, without rigorous statistical analysis, can lead to misinterpretation of the data in the situations typified by the EU referendum, where multiple factors influence a decision simultaneously. In such cases, in order to draw more meaningful conclusions advanced statistical analysis techniques are required. This paper, therefore, provides the first detailed assessment of the election results; in doing so we adopt both Multivariate Regression and Logit Model techniques, coupled with Tableau Public Tool to visualize our main findings, determining the dominant factors that have influenced the Brexit outcome.

The remainder of the article is organised as follows: Section 2 provides the details of the data collections, the research methodologies, as well as the data analysis and then main outputs

¹ The 2011 census is the extant definitive census of the UK population.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

using both Multivariate Regressions and the Logit Model. The key findings within the analysis are concluded in Section 3, where some commentary on policy implications is also made.

2. Research Design

To capture the joint impact of various potential demographic factors on the UK's decision to leave the EU, we use two separate methodologies to determine the significant factors.

First, multivariate regressions using Stata are carried out to identify statistically significant factors that have influenced the voting preference. We then use a Logit Model to confirm such findings as well as the odd ratios in favour of Leave. Before that we need to collect valid and confirmed datasets from reputable sources.

2.1 Data Collection

For these assessments the data have been obtained from: The Electoral Commission (2016); The Office for National Statistics (ONS) (2011); National Records of Scotland (2011) and; Northern Ireland Statistics & Research Agency (2011).

The EU referendum results from the Electoral Commission contain the data for *Electorate* and voting return results, from which the following independent variable of *Turnout* is derived. And it is defined as:

- *Turnout* = Ratio of both Remain and Leave votes to Electorate.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

The referendum results cover 382 voting areas, including 348 local authority areas in England and Wales, 32 council areas in Scotland and 1 aggregate area in Northern Ireland as well as the British Overseas Territory of Gibraltar².

Average annual pay is obtained from the ONS for the independent variable of *Income* for all 381³ areas in the UK and is defined as.

- *Income* = Average annual pay (£000)

The 2011 census data derived from Nomis contains 348 local authority areas in England and Wales and furnishes the following independent variables:⁴

- *Male* = percentage of British males.
- *NoQ* = percentage of British with no qualifications.
- *Educ* = percentage of British with Level 4 qualification and above (i.e. higher education).
- *Unem* = percentage of unemployed British (excluding full-time students)⁵.

² As this research is focus on the UK, the voting area in Gibraltar is not considered, although it is counted for the purposes of overall turnout and Brexit vote as the MEP for SW England also represents Gibraltar in the European Parliament.

³ Among the 381 areas, income data for 7 areas are missing.

⁴ We use the population data of Britons aged 16 (unless otherwise stated) and above who hold a passport or with no passport held as a proxy, due to the fact that the census data does not divide age groups by age 18, a cut-off age for being eligible for voting. On the other hand, the aged 16+ population in 2011 have grown to aged 20+ by the EU referendum in 2016, thus providing a good approximation to the EU referendum electorate.

⁵ The unemployed excludes economically inactive people (that are the retired, students, people looking after home or family, long-term sick/disabled).

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

- *ABCI* = percentage of estimated Social Grades AB and C1 aged 16 to 64.
- *Elderly* = percentage of British aged 65 and above.
- *UKborn* = percentage of British born in the UK (as % of whole local population).⁶
- *Differ* = difference between electorate (in 2016) and total British aged 16+ (2011 census), as a percentage.

The 2011 census data for Scottish council areas and Northern Ireland have been obtained from the National Record of Scotland and Northern Ireland Statistics & Research Agency separately.

2.2 Multivariate Regression Analysis

As shown in Figure 1, all of the 32 Scottish council areas and Northern Ireland voted for Remain, and only England and Wales voted for Leave. The population densities are different (or lower to be more precise) in Scotland and Northern Ireland than those in England and Wales. So the voting patterns may be different across these two groups. Bearing this in mind, we have carried out the multivariate regression analysis for these two groups separately. In the first instance we have analysed the data for England and Wales before moving to Scotland and Northern Ireland.

⁶ This factor is different from other percentage variables, which use British aged 16+ population as a base. This variable, *UKborn*, uses the entire population of all age groups and all nationals as a base. This can be justified as the effect of overseas born population on UK voters' opinion, in individual areas, is independent of age and nationality.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Figure 1: Leave, as % of total votes in each voting area

Sources: produced by the author from the referendum voting data. The heat map produced using Tableau Public is available at

https://public.tableau.com/views/2_1_10/2_1RemainvLeave?:embed=y&:display_count=yes

2.2.1 England and Wales

In the multivariate regression model, the dependent variable is chosen as the proportion of Leave votes across areas and is defined by *Leave* as

- *Leave* = the ratio of Leave votes to total votes.

We have assumed that the dependant variable, *Leave*, depends on specific local area characteristics (as reported by public media such as The Guardian, 2016), including place of birth, age, income, sex, social grades, education, unemployment and turnout. In addition, we also take into account the change in British adult population over the period of 2011 (when last census was conducted) and 2016 (when the plebiscite took place); this is captured by the variable *Differ*. The multivariate regression model then reads:

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

$$Leave_i = \alpha + \beta_1 Differ_i + \beta_2 Turnout_i + \beta_3 Male_i + \beta_4 NoQ_i + \beta_5 Educ_i + \beta_6 Unem_i + \beta_7 Elderly_i + \beta_8 ABC1_i + \beta_9 UKborn_i + \beta_{10} Income_i + \varepsilon_i \quad (1)$$

Where the subscript i stands for each voting area; ε is the error term; the coefficient α is the intercept (or constant); while each β_i measures the sensitivity of the dependent variable *Leave* to the corresponding independent variable. The coefficients α and β_i s are to be estimated from the actual data we collect (see 2.1 Data Collection).

To estimate the coefficients, we run various regressions using ordinary least squares (OLS) in Stata. As some of the 2011 census data is provided separately for England and Wales, Scotland and Northern Ireland, we first run the regressions for England and Wales to identify the key significant factors. We then focus on the identified significant factors and run the reduced regressions for Scotland and Northern Ireland. It is also reasonable to consider England and Wales separately from Scotland and Northern Ireland when examining the voting patterns, as the population densities in Scotland and Northern Ireland are significantly lower than those in England and Wales. Indeed, the estimated coefficients do differ across the two groups as presented below (compare estimated equations (2) and (4)).

As can be seen in

Table 1, the factors *Unem*, *UKborn* and *Income* appear to be insignificant. However, the correlation matrix in Table 2 shows high correlations (greater than 0.75) between *Unem* with *Turnout*, *ABC1* with both *NoQ* and *Educ*, and *Educ* with *NoQ*. A negative high correlation of -0.85 between *Turnout* and *Unem* implies that areas with a lower unemployment rate tend to have a higher turnout (see Figure 2). This phenomenon supports Radcliff's argument that a poor economy discourages voters' participation by those most adversely affected (see Radcliff, 1992). The high correlations of *ABC1* and *Educ* are obvious because the social grades are estimated from the information collected in the 2011 Census and based primarily

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

on their occupation which also includes information about their qualifications gained etc.

Thus a higher percentage of higher education qualification is associated with a higher percentage of social grades ABC1 (a correlation of 0.87) and a lower percentage of no qualifications (a correlation of -0.87). Areas with a higher proportion of university educated people have a lower proportion of people with no qualifications, thus a negative correlation of -0.87 between *Educ* and *NoQ*. All these show problems of multicollinearity in the regression model (1). To remedy such problems, we remove those highly correlated independent variables (*Unem*, *NoQ* and *ABC1*) as well as the insignificant variables (*UKborn* and *Income*)⁷ and rerun the regression, the results are presented in

Table 3.

Table 1: Regression result for England and Wales

Independent Variables	Estimated coefficient β	t-statistics	p-value for testing $\beta=0$	R-squared	p-value for testing R-squared = 0
Differ	0.21	3.96	0.00 (***)	0.87	0.00 (***)
Turnout	0.71	8.7	0.00 (***)		
Male	1.36	4.8	0.00 (***)		
NoQ	0.35	2.95	0.00 (***)		
Educ	-1.32	-19.51	0.00 (***)		
Unem	0.39	1.07	0.28		
Elderly	0.29	3.19	0.00 (***)		
ABC1	0.16	1.96	0.051 (*)		
UKborn	-0.01	-0.34	0.74		
Income	0.001	1.11	0.27		
intercept α	-0.53	-3.00	0.00 (***)		

Note: the stars in parentheses indicate statistical significance at 0.1 level (*), 0.05 level (**) and 0.01 (***) level, respectively.

⁷ Several regressions have been run in background to check the insignificant variables after removing the multicollinearity problem.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Table 2: Correlation Matrix for the Independent Variables in Eq. (1)

Correlation	Differ	Turnout	Male	NoQ	Educ	Unem	Elderly	ABC1	Ukborn	Income
Differ	1.00									
Turnout	0.15	1.00								
Male	0.17	-0.20	1.00							
NoQ	-0.55	-0.50	-0.24	1.00						
Educ	0.59	0.31	0.30	-0.87	1.00					
Unem	-0.12	-0.85	0.12	0.53	-0.33	1.00				
Elderly	-0.30	0.52	-0.38	0.07	-0.13	-0.59	1.00			
ABC1	0.62	0.30	0.29	-0.87	0.87	-0.28	-0.37	1.00		
UKborn	-0.28	0.18	-0.21	0.23	-0.29	-0.17	0.36	-0.35	1.00	
Income	0.51	-0.09	0.52	-0.46	0.57	0.08	-0.43	0.61	-0.27	1.00

Figure 2: Turnout vs Unemployment in England and Wales

Sources: produced by the author from the referendum voting data and 2011 census data. The

figure produced using Tableau Public is available at

https://public.tableau.com/views/2_2_8/2TurnoutvUnem?:embed=y&:display_count=yes

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Table 3: Regression result with a constant for England and Wales

Independent Variables	Estimated coefficient β	t-statistics	p-value for testing $\beta=0$	R-squared	p-value for testing R-squared = 0
Differ	0.22	4.31	0.00 (***)	0.86	0.00 (***)
Turnout	0.59	10.84	0.00 (***)		
Male	1.18	4.91	0.00 (***)		
Educ	-1.35	-38.46	0.00 (***)		
Elderly	0.16	2.55	0.011 (**)		
intercept α	-0.14	-1.11	0.27		

Table 4: Regression result without a constant for England and Wales

Independent Variables	Estimated coefficient β	t-statistics	p-value for testing $\beta=0$	R-squared	p-value for testing R-squared = 0
Differ	0.22	4.26	0.00 (***)	0.996	0.00 (***)
Turnout	0.57	11.21	0.00 (***)		
Male	0.92	15.11	0.00 (***)		
Educ	-1.34	-40.27	0.00 (***)		
Elderly	0.15	2.46	0.014 (**)		

As can be seen in

Table 3, the intercept α now becomes insignificant while the overall explanatory power ($R^2 = 0.86$) of the model remains almost the same as that in Table 1, indicating that the independent variables (*Differ*, *Turnout*, *Male*, *Educ*, and *Elderly*) jointly account for 86% of the variability in the referendum vote across areas in England and Wales. For comparison, let us further drop the insignificant constant α and then rerun the regression. The result is presented in Table 4, from which we see that all the independent variables remain statistically significant and that all the estimated coefficients are similar to those reported in Table 3, except the coefficient for *Male*. However, the R^2 becomes nearly 100%, which could be

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

considered artificial due to the regression without a constant.⁸ Keeping this in mind, we therefore use the regression model with a constant to study the marginal effects of the significant variables (factors) on the referendum result by examining each of the estimated coefficients β_i s in

Table 3. The fitted equation of the regression with a constant is given by

$$\widehat{Leave}_i^9 = -0.14 + 0.22 \times Differ_i + 0.59 \times Turnout_i + 1.18 \times Male_i - 1.35 \times Educ_i + 0.16 \times Elderly_i \quad (2)$$

First, we can see that all factors, except *Educ*, have a positive sign, implying that there is a positive relationship between each of them and the dependent variable *Leave*. For example, a higher percentage of turnout drives the proportion of Leave votes higher across areas. And the marginal effects are given by the estimated numbers in absolute value. Thus a coefficient of 0.59 indicates that one extra percentage of turnout can push the percentage of Leave votes in an area higher by 0.59%, keeping all other factors unchanged. Similarly, keeping other things equal, an extra percentage of male population in an area can push the percentage of its Leave votes higher by 1.18%; an extra percentage of elderly population (aged 65+) in an area can push the Leave votes higher by 0.16%; a 1% increase of British adults, who now reside in an area, can push the area's Leave result higher by 0.22%, this increase being explained by either inward migration (over the last 5 years) and/or those British teenagers (aged 13-18 in 2011) who have grown eligible for the referendum vote, by the passage of time since last

⁸ An interesting discussion about why R^2 are so large for models without a constant is discussed here

http://www.ats.ucla.edu/stat/mult_pkg/faq/general/noconstant.htm.

⁹ The hat (^) sign indicates the estimated value of *Leave*.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

census¹⁰. The university educated people tend to vote for Remain and an extra percentage of university educated people in an area will push the Remain votes higher by 1.35% (due to the negative coefficient of -1.35), keeping other things equal. This negative relationship between *Educ* and *Leave* is clearly shown in Figure 3. The Boxplot in Figure 4 further shows that most of the Remain areas have a higher proportion of British population with higher education than the national average of 26% (see also Figure 5) while about half of the Leave areas below this national average. And about half of the Remain areas have a higher percentage of British population with higher education than the highest percentage of 36% of the Leave areas.

It is also interesting to note that, by simple calculations using (2), a decrease of about 7%¹¹ in turnout in England and Wales could reduce the Leave votes by circa 4%. This could result in a Remain victory in the UK.¹² Similarly, an about 3% increase in the proportion of British adults accessing to higher education in England and Wales could also reverse the referendum result. Predominantly, the higher education factor alone explains about 77% of the total variation in the referendum vote across areas in England and Wales.¹³

¹⁰ Recall that *Differ* is defined as the percentage difference in the number of British who are eligible for the Referendum vote in 2016 and the number of British aged 16+ recorded in 2011 census.

¹¹ A 7% decrease in turnout in England and Wales could bring the overall turnout in the UK back to the 2015 general election turnout of about 66%.

¹² A decrease in the Leave votes by 4% in England and Wales reduces the total Leave votes in England and Wales by 641,686 of the total Leave votes of 16,042,155 in these two Brexit regions. This reduction of Leave votes counts for 2% of the total Leave votes in the UK. Adding this 2% to the Remain votes will result in Remain winning by about 0.1% (because the referendum result is 51.9% / 48.1% for Leave/Remain).

¹³ The simple regression is run in the background but not presented here for the sake of being concise.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Figure 3: Higher Education vs Leave in England and Wales, UK mean 26%

Sources: produced by the author from the referendum voting data and 2011 census data. The figure produced using Tableau Public is available at

https://public.tableau.com/views/2_3_6/2_3ScatterHE?:embed=y&:display_count=yes

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Figure 4: Boxplot of Higher Education in England and Wales, UK mean 26%

Sources: produced by the author from the referendum voting data and 2011 census data. The figure produced using Tableau Public is available at

https://public.tableau.com/views/2_4_2/2_4BoxplotHE?:embed=y&:display_count=yes

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Figure 5: Higher Education in England and Wales, as %, UK mean 26%

Sources: produced by the author from the referendum voting data and 2011 census data. The heat map produced using Tableau Public is available at

https://public.tableau.com/views/2_5_1/2_5HE?:embed=y&:display_count=yes

2.2.2 Scotland and Northern Ireland

For the analysis of Scotland and Northern Ireland, we add a subscript ‘ $_S$ ’ to each variable to denote the corresponding variable for Scotland and Northern Ireland. Following the arguments for England and Wales above, we restrict ourselves on the significant independent variables only, namely, *Differs*, *Turnouts*, *Males*, *Eldderys* and *Educ*, and then estimate the coefficients β_i s using OLS as before.

$$\begin{aligned}
 Leave_{S_i} = & \alpha_i + \beta_1 \times Differ_{S_i} + \beta_2 \times Turnout_{S_i} + \beta_3 \times Male_{S_i} + \beta_4 \times Educ_{S_i} \\
 & + \beta_5 \times Elderly_{S_i} + \varepsilon_i
 \end{aligned} \tag{3}$$

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

The estimation result of (3) is summarized in Table 5. Variables *Differ_S* and *Turnout_S* are insignificant while the overall explanatory power measured by R^2 is 76%. From the correlation matrix Table 6, we see that *Turnout_S* is highly correlated with *Educ_S* with a correlation coefficient of 0.74, indicating that areas in Scotland and Northern Ireland with a higher proportion of university educated people tend to have a higher turnout to support Remain (see Figure 6).¹⁴

Table 5: Regression result for Scotland and Northern Ireland

Independent Variables	Estimated coefficient β	t-statistics	p-value for testing $\beta=0$	R-squared	p-value for testing R-squared = 0
Differ_S	-0.21	-1.46	0.16	0.76	0.00 (***)
Turnout_S	0.07	0.31	0.76		
Male_S	3.89	5.8	0.00 (***)		
Educ_S	-0.73	-4.17	0.00 (***)		
Elderly_S	0.55	2.2	0.04 (**)		
intercept α	-1.47	-4.53	0.00 (***)		

Table 6: Correlation Matrix for the Independent Variables in Table 5

Correlation	Differ_S	Turnout_S	Male_S	Educ_S	Elderly_S
Differ_S	1.00				
Turnout_S	0.16	1.00			
Male_S	-0.07	0.16	1.00		
Educ_S	0.41	0.74	0.13	1.00	
Elderly_S	-0.05	0.58	0.07	0.28	1.00

¹⁴ Note that Northern Ireland is not shown on Figure 6 due to the aggregated electorate size of it.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

Figure 6: Turnout vs Higher Education in Scotland

Sources: produced by the author from the referendum voting data and 2011 census data. The figure produced using Tableau Public is available at

https://public.tableau.com/views/2_6_3/2_6TurnoutvHE?:embed=y&:display_count=yes

The final result after removing the insignificant variables is presented in Table 7 and Equation (4). As in England and Wales, areas in Scotland and Northern Ireland with a higher proportion of university educated people tend to vote for Remain, but with a smaller marginal effect. An extra percentage increase of the British population with higher education in Scotland and Northern Ireland would reduce Leave votes by 0.77%, whereas in England and Wales the reduction would be 1.35%. This is not surprising as the proportion of Leave votes

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

in England and Wales is much higher than that in Scotland and Northern Ireland. As in the case of England and Wales, both higher proportions of British males and elderly in local areas tend to push their Leave votes higher in Scotland and Northern Ireland (note the positive signs of the estimated coefficients for both factors), but with much bigger marginal effects. In particular, the marginal effect of *Male* in Scotland and Northern Ireland on the referendum result is more than three times that in England and Wales.

Table 7: Regression result for Scotland and Northern Ireland

Independent Variables	Estimated coefficient β	t-statistics	p-value for testing $\beta=0$	R-squared	p-value for testing R-squared = 0
Male_S	4.04	6.05	0.00 (***)	0.74	0.00 (***)
Educ_S	-0.77	-7.08	0.00 (***)		
Elderly_S	0.65	3.16	0.00 (***)		
Constant	-1.5	-4.7	0.00 (***)		

$$\widehat{Leave}_{S_i} = -1.5 + 4.04 \times Male_{S_i} - 0.77 \times Educ_{S_i} + 0.65 \times Elderly_{S_i} \quad (4)$$

2.3 The Logit Model

In the case where individuals are making a choice between Leave and Remain, we can formalise the situation by specifying a utility function (e.g. a log utility function)¹⁵ to say that the individuals will make the choice generating a higher utility (or higher satisfaction).

Let D_i denote the difference in overall utility (or net utility) that individuals in area i obtained between voting Leave and Remain. The area i votes for Leave (overall) if the utility

¹⁵ In economics, a utility function is a real number – valued function, which measures preferences over some set of goods (including services: something that satisfies human wants); it represents satisfaction experienced by the consumer of a good.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

difference is positive (that is $D_i > 0$), and vote for Remain otherwise.¹⁶ Note that the utility across the two choices is unobservable. However, we can observe the choices that individuals in each voting area have actually made, thanks to the electoral data collected by the Electoral Commission. Now we can define a binary variable Y and relating it to D by the following equations for each area:

$$Y_i = 1 \text{ if } D_i > 0$$

$$Y_i = 0 \text{ if } D_i \leq 0,$$

Where, i denotes area i .

Thus $Y_i = 1$ if the number of Leave votes is greater than the number of remain votes in area i (and 0 otherwise). Let p_i denote the probability that area i votes for Leave, then the probability it votes for Remain is $1 - p_i$. We know that area i will vote for Leave if the utility difference is positive. So we have the following equation¹⁷

$$p_i = P(Y_i = 1) = P(D_i > 0).$$

Taking expectation of Y_i , we get the average value of Y_i as

$$E(Y_i) = p_i.$$

The logit model reads as¹⁸

¹⁶ An area votes for Leave if the proportion of Leave votes in that area is greater than 50%, otherwise it votes for Remain.

¹⁷ In the logit model, we only focus on the significant variables identified in the multivariate regression analysis.

¹⁸ A probit model can also be applied here, but as argued by Koop (2013) and Gujarati (2011)), probit and logit models tend to yield very similar results.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha_i + \beta_1 \times \text{Differ}_i + \beta_2 \times \text{Turnout}_i + \beta_3 \times \text{Male}_i + \beta_4 \times \text{Educ}_i + \beta_5 \times \text{Elderly}_i + \varepsilon_i \quad (5)$$

The dependent variable of (5) is the natural log of the odds ratio in favour of Leave.

For the logit model, we first apply it to the whole UK (that is, England, Wales, Scotland and Northern Ireland) data. We then compare the results with those of England and Wales.

The results from the logit model for the UK produced in Stata are summarized in Table 8 and Table 9, and the results for England and Wales are summarized in Table 10 for comparison.

Table 8: Results from the Logit Model for the whole UK

Independent Variables	Logit coefficient β	p-value for testing $\beta=0$	Marginal effect on probability of Leave	Pseudo R-squared	E(Y) = P(Leave) (predict)
Differ	13.69	0.00 (***)	2.02	0.58	0.82
Turnout	53.92	0.00 (***)	7.94		
Male	130.59	0.00 (***)	19.23		
Educ	-71.26	0.00 (***)	-10.49		
Elderly	3.44	0.61	0.51		
constant	-83.48	0.00 (***)			

From Table 8, we can see that the variable *Elderly* becomes insignificant while the other variables remain significant and with the same signs as identified in the multivariate regression analysis. Note also that this factor *Elderly* was consistently less significant than the other significant variables in the multivariate regression analysis above (see

Table 3, Table 4, Table 5 and Table 7). So, combining the results from both the multivariate regression analysis and the logit model, we can conclude that a higher proportion of elderly British can contribute to a relatively higher percentage of Leave votes (a result from the multivariate regressions), however, these extra Leave votes do not lead to a Leave outcome in individual areas on their own (a result from the logit model). We then remove the

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

insignificant variable from (5). This results Table 9, in which all remaining factors are highly significant.

Table 9: Results from the Logit Model for the whole UK

Independent Variables	Logit coefficient β	p-value for testing $\beta=0$	Marginal effect on probability of Leave	Pseudo R-squared	E(Y) = P(Leave) (predict)
Differ	12.82	0.00 (***)	1.88	0.58	0.82
Turnout	55.64	0.00 (***)	8.17		
Male	126.98	0.00 (***)	18.65		
Educ	-71.43	0.00 (***)	-10.49		
constant	-82.15	0.00 (***)			

The estimated odds ratio in favour of Leave in the UK is then given by¹⁹

$$\frac{\hat{p}_i}{1-\hat{p}_i} = \text{Exp}(-82.15 + 12.82 \times \text{Differ}_i + 55.64 \times \text{Turnout}_i + 126.98 \times \text{Male}_i - 71.43 \times \text{Educ}_i) \quad (6)$$

From the results in Table 9, we can answer “How much does the probability of voting for Leave change when we alter each of the independent variables?” The answer is given by the “marginal effect on probability of Leave” in the 4th column of Table 9. While the demographic changes in British adult population over the period of 2011 and 2016 (captured by *Differ*) and sex structure in Britain cannot normally be altered (both factors of *Differ* and *Male* are statistically highly significant as in the multivariate regression case though), the turnout and education could be influenced. Decreasing turnout by 1% in an area can decrease the probability of Leave by 8.17% in the UK and 3.05% in England and Wales, keeping other factors unchanged. Similarly, increasing the proportion of British adults who receive a university degree and above by 1% can decrease the probability of Leave by 10.49% in the UK and 5.31% in England and Wales, keeping other things equal, while the marginal effect

¹⁹ The hat (^) sign indicates the estimated value and *Exp* is the exponential function.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

of higher education on the probability of Leave outcome remains the strongest among other factors in England and the Wales. Given the average percentage change in British adult population over last 5 years (captured by *differ*) of -1.8%, the average turnout of 73.6%²⁰, the average percentage of British male adults of 48.7% and the average percentage of university educated British adults of 26.4%, the predicted probability of voting for Leave in the UK is about 0.82, giving the odds ratio in favour of a Leave of about 5:1 ($=0.82/(1-0.82)$).²¹

Table 10: Results from the Logit Model for England and Wales

Independent Variables	Logit coefficient β	p-value for testing $\beta=0$	Marginal effect on probability of Leave	Pseudo R-squared	E(Y) = P(Leave) (predict)
Differ	20.88	0.00 (***)	1.22	0.68	0.94
Turnout	52.04	0.00 (***)	3.05		
Male	70.32	0.05 (**)	4.12		
Educ	-90.49	0.00 (***)	-5.31		
constant	-45.71	0.00 (***)			

The estimated odds ratio in favour of Leave in England and Wales is then given by

$$\frac{\hat{p}_i}{1-\hat{p}_i} = \text{Exp}(-45.71 + 20.88 \times \text{Differ}_i + 52.04 \times \text{Turnout}_i + 70.32 \times \text{Male}_i - 90.49 \times \text{Educ}_i) \quad (7)$$

In England and Wales, where most Leave voters reside, the predicted probability of voting for Leave is about 0.94 (see Table 10), giving the odds ratio in favour of Leave of about 16:1

²⁰ Note that the average turnout is calculated as the mean of the turnouts across areas, thus different from the overall turnout of 72.2%.

²¹ The odds ratio can also be calculated using (6). However, caution should be taken when using the estimated coefficients as they are rounded numbers and very small numbers can result in significantly large differences in the odds ratio. It is therefore better to use the predicted probability calculated by Stata itself as quoted in Table 9.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

($=0.94/(1-0.94)$), about 3 times as the odds ratio for the entire UK. The average values for *Differ*, *Turnout*, *Male* and *Educ* across areas in England and Wales are -1.7%, 74.2%, 48.7% and 26.6%, respectively.

3. Conclusion and Discussion

3.1 Conclusion

On June 23, 2016, Britain's vote to Leave the EU came as a surprise to most observers, with a bigger voter turnout (72.2%) than that of any UK general election in the past decade²². On morning of the referendum, The Economist's DATA TEAM (2016b) claimed that a high overall turnout would likely benefit Remain, as younger and richer elderly voters are more supportive of being a member of the EU and this population is normally the highest proportion of the electorate who vote. The statistical analysis of the actual data, conducted in this paper and including the actual EU referendum results provided by the Electoral Commission (2016) and the 2011 census data from England and Wales, Scotland and Northern Ireland, reveals rather different conclusions.

The effect of voter turnout is one of the significant factors influencing the referendum result. The high overall turnout actually benefited Leave with a big proportion of the turnout being from enthusiastic leave supporters, with a corresponding under turnout of remain supporters. This could indicate that the Leave campaign was more effective in mobilising their support base, but may also explain that regardless of the effectiveness of the campaigns the Leave voters felt more strongly about the factors affecting Leave than did the Remain supporters.

²² The turnouts of UK general elections were 66.4% in 2015, 65.1% in 2010 and 61.4% in 2005, respectively (see The Electoral Commission, 2016).

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

We also find that areas in England and Wales with a lower unemployment rate tend to have a higher turnout to support Leave while areas in Scotland and Northern Ireland with a higher proportion of university educated people have a higher turnout to support Remain. As previously discussed, a decrease of about 7% in turnout in England and Wales could reduce the Leave votes by circa 4%, which could have resulted in a Remain victory in the UK.

Moreover, higher education is found to be the predominant factor dividing the nation, in particular in England and Wales, between Remain and Leave. This analysis demonstrates highly significant evidence that university educated British people tend to vote consistently across the UK for Remain. It can be speculated that, as much of the Leave campaign, was characterised by emphasising detrimental factors (such as immigration or the recession of the economy, see The Economist, 2016) the university educated voters seem to be more immune than those who do not have university education, to this kind of campaign. The marginal effect of higher education on the referendum decision is much stronger than any other factors in England and Wales, where most Leave voters reside. As demonstrated in Section 2.2.1, if a 3% increase in the proportion of British adults accessing to higher education in England and Wales were to be realised, this could have reversed the referendum result.

We also find highly significant evidence that demographic changes of British adult males in individual areas, have an effect on the referendum outcome with areas with a higher proportion of British male adults being associated to a higher percentage of Leave votes.

The impacts of the proportion of UK born British and income are found to be statistically insignificant.

The factor of elderly voters, although having an effect on the outcome, was generally over claimed as a dominant factor. Combining the findings from both the multivariate regression analysis and the logit model, we conclude that a higher proportion of elderly British can

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

contribute to a relatively higher percentage of Leave votes (a result from the multivariate regressions), however, these extra Leave votes do not lead to a Leave outcome in individual areas on their own (a result from the logit model).

The logit model also allows us to predict the odds ratio in favour of Leave. We find that the odds ratio in favour of Leave in England and Wales is about 3 times as the odds ratio for the entire UK.

3.2 Discussion

The above findings may have the following policy implications:

It can be implied from the analytical results above that a lack of further or higher education affects a population's attitude to change, which could be assigned to, paradoxically, feelings of and inability to affect political changes, caused by a perceived reduced access to a highly educated 'political elite'. This could have a further consequence in future elections where the results become less predictable due to broad public sentiment over established political dogma.

This leads to a conclusion that a greater access to higher and further education would produce different political outcomes, which has been potentially demonstrated in the 2017 General Election, where it can be argued that the elements of the voting population with a higher education (especially in the younger voting population) had a decisive effect on the result (see Shrimley, 2017).

The research also indicates that the Leave campaign was effective in motivating its supporters to turnout to vote. Those campaigning to remain put their money on the polls being right. All indications from the polls were that Britain would remain in the EU. So the campaign vigor for those wanting to remain abated in comparison with those seeking to exit

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

(Simmons, 2016). Some of the voters who supported Leave might have been less engaged in the general elections than in the EU referendum; this was potentially combined with the perceived 'forgone' conclusion that Remain had of victory; with a consequential lower perceived need for the Remain voters to turnout. So this indicates that hubris can have a significant effect on the referendum outcomes.

Acknowledgements

The author would like to thank the following people (in no particular order) who have helped her with obtaining various data sets required for the analysis: CAZ (Nomis Census Customer Services), Matt Singh (Number Cruncher Politics), Ben Davoll (Special EU Programmes Body), DoF Census NISRA, Pawel Kosinski (Customer Services of National Records of Scotland). Great thanks go to Matthew Blake for various useful suggestions and inputs as well as proof reading the draft. I would also like to thank my colleagues (Clive Rix, Jeremy Levesley, Leena Sodha, Frank Neumann and Alexander Gorban and Christian Ewald) for their gracious support. The author would also like to thank the anonymous reviewers for their very supportive comments and suggestions.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

References

BBC News (2016). EU referendum: the result in maps and charts, June 24. Available at <http://www.bbc.co.uk/news/uk-politics-36616028> accessed on Nov 22, 2016.

The Data Team (2016a). Britain votes to leave the EU. *The Economist*, June 24. Available at <http://www.economist.com/blogs/graphicdetail/2016/06/daily-chart-17> accessed on Nov 21, 2016.

The Data Team (2016b). The Economist's "Brexit" poll-tracker. *The Economist*, June 23. Available at <http://www.economist.com/blogs/graphicdetail/2016/06/britain-s-eu-referendum> accessed on Nov 21, 2016.

The Economist (2016). The referendum campaign: The Battle of Evermore. *The Economist*, June 18, 2016.

The Electoral Commission (2016). Electoral data. Available at <http://www.electoralcommission.org.uk/our-work/our-research/electoral-data> accessed on Nov 22, 2016

Gujarati, D. (2011). *Econometrics by Example*. Palgrave Macmillan.

The Guardian (2016). EU referendum: full results and analysis. Available at <https://www.theguardian.com/politics/ng-interactive/2016/jun/23/eu-referendum-live-results-and-analysis> accessed on Nov 22, 2016.

Kirk, A. (2016). EU referendum: Which type of person wants to leave, and who will be voting to remain? *The Telegraph*, June 22. Available at <http://www.telegraph.co.uk/news/2016/06/22/eu-referendum-which-type-of-person-wants-to-leave-and-who-will-b/> accessed on Nov 21, 2016.

Koop, G. (2013). *Analysis of Economic Data*, 4th Edition, Willey.

KEY FACTORS INFLUENCING THE UK'S REFERENDUM

National Record of Scotland (2011). Scotland's census 2011. Available at <http://www.scotlandscensus.gov.uk/ods-web/standard-outputs.html> accessed on Nov 18, 2016.

Northern Ireland Statistics & Research Agency (NISRA) (2011) Census. Available at <http://www.nisra.gov.uk/census/2011Census.html> access on Nov 18, 2018.

Office for National Statistics (2011). Census Data on Nomis www.nomisweb.co.uk accessed Nov 18, 2016.

Radcliff, B. (1992). The Welfare State, Turnout, and the Economy: A Comparative Analysis. *American Political Science Review* 86:444-54.

Simmons, A.M. (2016). The British establishment didn't think 'Brexit' could win, but it did. Here's why. *Los Angeles Times*, June 24. Available at <http://www.latimes.com/world/europe/la-fg-brexit-what-went-wrong-20160624-snap-story.html>, accessed on July 4, 2017.

Shrimpsley, R. (2017). Six myths of the UK election. *Financial Times*, June 12. Available at <https://www.ft.com/content/283dd902-4f59-11e7-a1f2-db19572361bb?segmentId=ac9e49a7-a213-db59-61c5-053662b6ec9c>, accessed on July 3, 2017.

Stolz, B., Harrington, H.A. and Porter, M.A. (2016). The Topological Shape of Brexit, October 30, Cornell University Library. Available at <https://arxiv.org/abs/1610.00752>