

Educational attainment alone can correctly classify over 90% of local authorities by voting outcome in the EU referendum



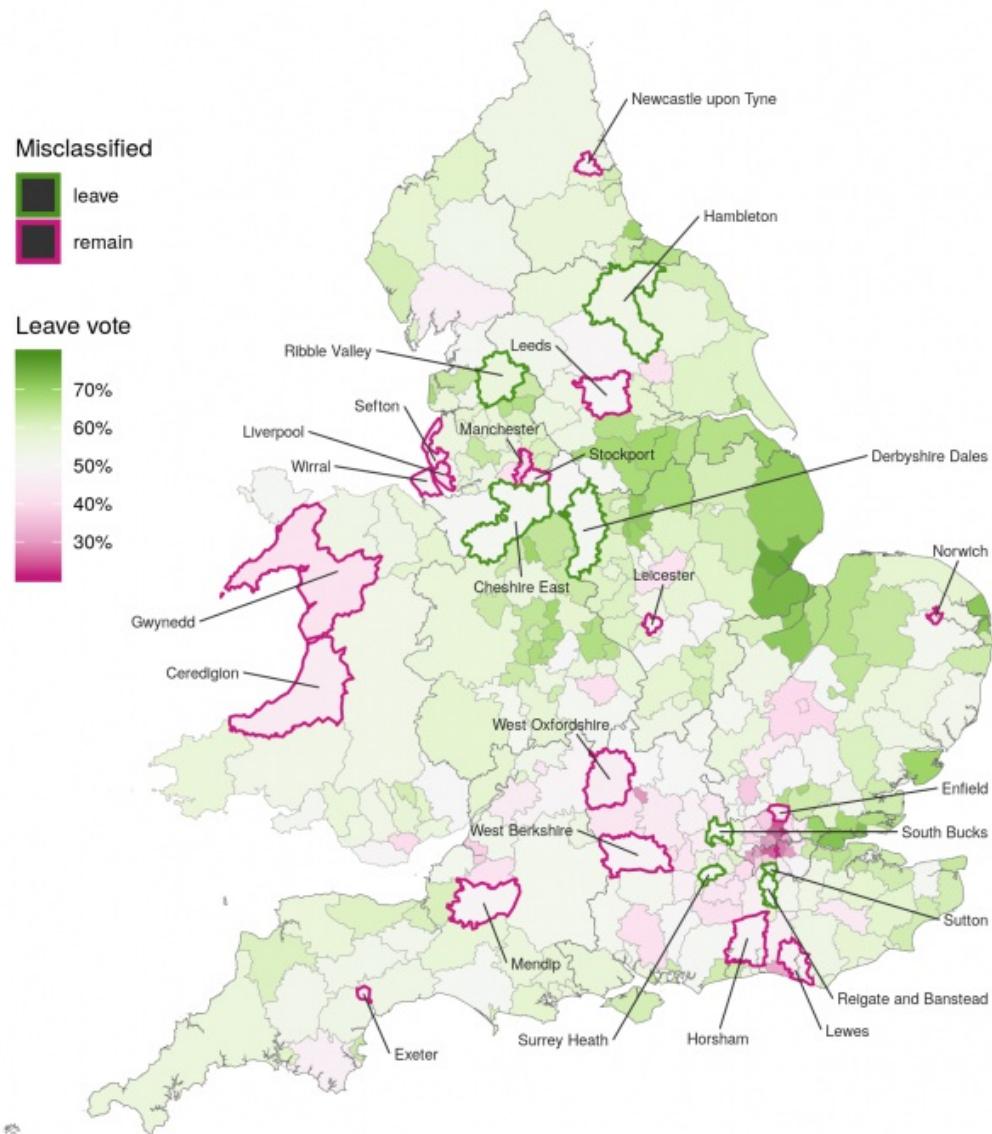
While it is well established that educational attainment is highly correlated with Brexit voting patterns, the predictive capacity of education has attracted less attention. Using full-sample and split-sample exercises, [Rob Calvert Jump](#) and [Jo Michell](#) demonstrate that educational attainment alone can correctly classify over 90% of local authorities by voting outcome in the 2016 referendum to leave the EU, depending on the prediction model and classification method used. This illustrates the importance of education as a key factor in the geography of Brexit.

In 2017, on the BBC's Sunday Politics programme, the Labour MP for Huddersfield argued that 'The truth is that when you look at who voted to Remain, most of them were the better-educated people in our country.' This observation provoked widespread debate, with [at least one Tory MP](#) claiming to be 'astounded by this snobbery'.

In a [recent paper](#) published in the *Journal of Elections, Public Opinion and Parties*, we examine the predictive link between educational attainment and Leave voting in the 2016 referendum: we classify voting districts as 'Leave' or 'Remain' based on the proportion of their population educated to degree-level, both with and without a set of demographic controls. Comparing our classifications with the actual referendum results provides a simple measure of the capacity of educational attainment to predict local authority referendum results. This is a common technique in statistics and data science (a simple description can be found [here](#)).

Using probit models, we find that educational attainment can correctly classify over 90% of voting districts in England and Wales, with success slightly higher for Leave areas. For example, a standard probit model of educational attainment, 'trained' on a random two-thirds of voting districts in England and Wales, correctly classifies 91% of the remaining districts on average. Interestingly, the addition of extra controls, including population shares born in the UK, identifying as ethnically white, identifying as male, median age, and a measure of average socio-economic status, only improves on the classification success of education by one or two percentage points. Using these variables *without* education, on the other hand, can reduce classification success by a much greater amount.

Figure 1 illustrates this result using a choropleth (a map in which data is displayed using a colour scale). Leave areas are shaded green, and Remain areas shaded pink, with the areas *misclassified* by a random effects probit model of educational attainment highlighted and labelled. There are only 26 misclassified areas out of a total of 348 voting districts in England and Wales, giving an overall classification success of 92.5%.



Note: City of London removed from scatters as an outlier with 68% degree educated

The capacity of educational attainment to predict Leave and Remain areas is, therefore, quite remarkable. But what does this tell us about causality? It is [very difficult](#) to answer questions of the form, 'what were the causes of Leave voting in 2016?' using statistical inference. In this specific example, it is also difficult to answer questions of the form, 'if average educational attainment increased prior to 2016, would it have reduced Leave voting in 2016?', not least because the referendum was a one-off event.

The fundamental issue is that the observed correlation between educational attainment and Leave voting might contain no information on causality, and may simply reflect omitted variables. Nevertheless, as the [data visualisation pioneer](#) Edward Tufte observed, 'correlation is not causation, but it sure is a hint.' One would, therefore, be forgiven for thinking that the strength of the observed relationship between educational attainment and Leave voting tells us *something* about the underlying causes of Leave voting, even if it is not obvious what that something is.

To explore this further, we can apply a [simple tool](#) recently proposed by Emily Oster to the [comprehensive examination](#) of the correlates of Brexit voting published by Sascha Becker, Thiemo Fetzer and Dennis Novy in 2017. The idea behind Oster's tool is that any changes in the observed relationship between education and Leave voting after the inclusion of *observable* control variables ought to be similar to the changes that *would* occur if we could control for *unobservable* variables.

Of the potential correlates of Leave voting in 2016, demographic variables such as birthplace, ethnicity, and gender are largely observable, as are economic variables such as unemployment rates, the incidence of austerity, and so on. But other variables thought to be important in politics, such as liberal values or attitudes to immigration, might be much more difficult to measure and may well be correlated with both educational attainment and Leave voting.

Becker et al. report that an increase in the share of the population educated to degree-level or above (as of 2001) by one standard deviation (around 7.3 percentage points) decreased the expected share of a district's Leave vote in the 2016 referendum by just over 8 percentage points. Including a full set of demographic controls reduces this effect to just over 6 percentage points, and the effect is further reduced when a 'blocked variable selection' procedure chooses a 'best' model out of a (very) large number of potential covariates. The lowest effect recorded by Becker et al. is 4.7, i.e., the inclusion of observable controls reduces the estimated effect of educational attainment on Leave voting by around 40%.

Using the effects estimated by Becker et al. before and after the inclusion of *observable* controls, we can use Oster's tool to give us some idea of the effect that *would* be estimated if we could include the remaining set of *unobservable* controls. Applying the restricted estimator in section 3.2 of Oster (2019) to the lowest effect recorded by Becker et al. yields a bias-adjusted treatment effect of just over 3. This is lower than 4.7, but greater than zero, and suggests that unobserved confounders are unlikely to fully explain the relationship between educational attainment and Leave voting observed in 2016.

As Becker et al. point out, their analysis 'cannot possibly establish causality' – and neither can our results on the classification success of educational attainment. Nevertheless, the observed relationship between educational attainment and Leave voting is extremely strong, and appears to be robust to the inclusion of a host of observed covariates. The latter observation, in particular, suggests that the likelihood that unobserved confounders account for the *entirety* of the estimated relationship between education and Leave voting is rather low.

In other words, while the identification of a causal link between educational attainment and Leave voting is extremely difficult, academics have yet to rule out its existence.

Note: the above draws on the authors' [published work](#) in the *Journal of Elections, Public Opinion and Parties*.

About the Authors

Rob Calvert Jump is a Research Fellow at the University of Greenwich, based in the Institute of Political Economy, Governance, Finance and Accountability.



Jo Michell is Associate Professor of Economics at UWE Bristol.

Photo by [Christian Lue](#) on [Unsplash](#).