Brexit vote

SUMMARY

On 23 June 2016, the British electorate voted to leave the European Union (EU). We analyse vote and turnout shares across 380 local authority areas in the United Kingdom. We find that exposure to the EU in terms of immigration and trade provides relatively little explanatory power for the referendum vote. Instead, we find that fundamental characteristics of the voting population were key drivers of the Vote Leave share, in particular their education profiles, their historical dependence on manufacturing employment as well as low income and high unemployment. At the much finer level of wards within cities, we find that areas with deprivation in terms of education, income and employment were more likely to vote Leave. Our results indicate that a higher turnout of younger voters, who were more likely to vote Remain, would not have overturned the referendum result. We also compare our UK results to voting patterns for the far-right leader Marine Le Pen in the 2017 French presidential election. We find similar factors driving the French vote. An out-of-sample prediction of the French vote using UK estimates performs reasonably well.

JEL codes: D72, N44, R23, Z13

—Sascha O. Becker, Thiemo Fetzer and Dennis Novy

Economic Policy October 2017 Printed in Great Britain

Who voted for Brexit? A comprehensive district-level analysis

Sascha O. Becker, Thiemo Fetzer and Dennis Novy*

Centre for Competitive Advantage in the Global Economy (CAGE), and University of Warwick, and CEPR, CESifo, ifo, IZA and ROA; Centre for Competitive Advantage in the Global Economy (CAGE), University of Warwick and SERC; and Centre for Competitive Advantage in the Global Economy (CAGE), University of Warwick and CEPR, CESifo and CEP/LSE

1. INTRODUCTION

The United Kingdom's relationship with the European Union (EU) has always been a very special one. Not being a founding member, the United Kingdom only joined the European Economic Community (EEC), the precursor of the EU, in 1973. Merely two years later, the United Kingdom held its first in-out referendum. It produced a clear two-thirds majority to remain as a member. The United Kingdom has historically been a key supporter of several

*This paper was presented at the 65th Panel Meeting of *Economic Policy* in April 2017. We are grateful for constructive comments received by the panel, our discussants Vasso Ioannidou and Andrea Mattozzi, the editor and three referees. We also thank Andrew Oswald, Alan Taylor, seminar participants at Warwick and HM Treasury as well as numerous commentators on social media. We thank the ESRC initiative "The UK in a Changing Europe" for financial support. We thank Bart Los for sharing data with us. Eleonora Alabrese and Ivan Yotzov provided outstanding research assistance.

The Managing Editor in charge of this paper was Thorsten Beck.

Economic Policy October 2017 pp. 601–650 Printed in Great Britain

core features of today's EU such as the Single Market and EU Regional Policy. However, the United Kingdom never joined the Euro. It did not follow the route of the six founding members of the European project (and many other countries) to proceed towards ever closer union [see Guiso *et al.* (2016) for an analysis of Euro membership in the context of the European integration project].[1] Over the last two decades, the United Kingdom seemed to have grown increasingly lukewarm towards the EU. During the 2015 general election campaign, internal struggles within David Cameron's Conservative party led him to promise a referendum on EU membership. This referendum happened on 23 June 2016.

The UK referendum on EU membership is thought to have been a watershed moment in European integration and globalization more broadly. Although the outcome had been expected to be tight, in the days running up to the referendum bookmakers and pollsters predicted the Remain side to win. Many observers were left puzzled and keen to understand who voted for Leave. Various newspapers and blogs quickly produced correlations between selected variables and the referendum result, but no study has so far taken a comprehensive approach to attempting to understand the Brexit vote.[2] Our paper fills this gap by combining a multitude of geographically disaggregated data sources to carry out a comprehensive descriptive analysis of the socio-economic characteristics that correlate with the outcome of the 2016 referendum.

In particular, we study the EU referendum results in England, Wales and Scotland disaggregated across 'local authority areas' of the referendum (and across 107 wards within four cities) and relate them to fundamental socio-economic features of these areas.[3] The EU Referendum Act passed by Parliament in 2015 divided the United Kingdom into 382 official counting areas (which are the same as local authority areas), 327 of which are in England, 22 in Wales and 32 in Scotland.[4] There are on average roughly 122,000 eligible voters per local authority area. Data are not provided at the level of individual polling stations.

As covariates, we focus on socio-economic characteristics that can be broadly grouped into four categories: measures of an area's exposure to the EU; measures capturing (the quality of) public services provision and exposure to fiscal consolidation (austerity); demographic and human capital characteristics; and measures capturing the underlying economic structure of an area.

1 See Online Appendix A for a more detailed history of Britain's role in the EU.

2 For instance, see Burn-Murdoch (2016b) in the *Financial Times* as an example of various correlation plots.

3 An analysis of voting at the local authority area level does not necessarily reflect individual voting behaviour, a phenomenon called 'ecological fallacy'. We deliberately want to understand these regional voting patterns. But our analysis of within-city variation goes one step towards addressing worries about ecological fallacy because of the much finer level of geographical disaggregation.

4 We drop Northern Ireland because election results were only published for Northern Ireland as a whole. This makes Northern Ireland an outlier by being the largest 'local authority' by an order of magnitude. We also drop Gibraltar, a British Overseas Territory ceded to Britain in 1713 under the Treaty of Utrecht, where many covariates are missing. Thus, we end up with 380 voting observations at the local authority level.

We adopt a simple machine learning method to capture the subsets of variables from each group that best 'predict' the actual referendum result. We cannot possibly give a causal explanation of the referendum result because the election outcome is multi-causal and multi-faceted. Nevertheless, a systematic analysis across an exhaustive range of socio-economic characteristics can be helpful in directing future research efforts that aim at carefully identifying specific mechanisms. One might be able to single out an individual predictor such as immigration from Eastern Europe and try to establish causality from this specific factor for the Vote Leave share. But this would run counter to the aim of this paper, which is rather to focus on predictive power by pulling together various dimensions of the vote pattern.

Our results indicate that even very simple empirical models can explain significant amounts of variation in the Vote Leave share and achieve good predictive performance. Which characteristics have significant explanatory power for Leave support? Surprisingly and contrary to much of the political debate in the run-up to the election, we find that relatively little variation in the Vote Leave share is explained by measures of a local authority area's exposure to the EU (e.g., due to immigration and trade exposure). Neither is much variation explained by measures capturing the quality of public services and fiscal consolidation. Rather, a significant amount of the variation can be linked to variables that seem hardly malleable in the short run by political choices (variables such as educational attainment, demography and industry structure). We document that similar patterns hold when we explore data on the EU referendum result across 107 wards in four English cities – which to the best of our knowledge this paper is the first to exploit.

Our findings thus suggest that there is a disconnect between the key correlates of the vote outcome and the topics dominating the political debate in the run-up to the election. How can we reconcile this disconnect? The political debate centred on two issues: the fiscal burden of EU membership and the exposure to European immigration since the enlargement of the EU in 2004. Perhaps the UK budget contribution resonated so strongly with the British electorate because public services and benefits were under severe strain not least due to fiscal cuts. If we think of fiscal cuts and migration as political choice variables, we can explore the extent to which the powerful predictors capturing the underlying fundamentals (educational attainment, demography and industry structure) interact with these variables that saw significant change over the course of the last decade. Our results highlight that policy choices related to pressure from fiscal cuts and migration are linked to a higher Vote Leave share especially when socio-economic fundamentals are 'weak' (low incomes, high unemployment), and when the local population is less able to adapt to adverse shocks (due to low qualifications).

We stress that while our paper focuses on the *variation* of vote shares across local authority areas with respect to key variables such as immigration and education, we have less to say about the overall *level* of support for Vote Leave. Put differently, our paper focuses on slope coefficients, not intercepts. This is important because in order to get a sense of the absolute number of people who voted for or against Brexit, one would need

to refer to data on individuals and how they voted. To some extent, such information is available through polling data, for instance as provided by Ashcroft (2016). Such polls indicate that the typical Leave voter is white, middle class and lives in the South of England. The proportion of Leave voters that are in the lowest two social classes (D and E) is less than one-third (see Dorling, 2016).

We also carry out a back-of-the-envelope calculation regarding turnout. Young people voted overwhelmingly in favour of Remain but had a lower turnout than older age groups. We find that a higher turnout of young voters would have been very unlikely to result in a different referendum outcome, partly because their turnout was already elevated compared with previous UK-wide elections.

We also explore the role of some short-run factors such as heavy rainfall and flooding on the referendum day as well as train cancellations in the South East of England. While we document that these did have a reducing effect on turnout, the reduction does not seem to have affected the overall result: the Remain campaign would have still lost on a sunny day.

Lastly, we also compare our UK results to explain the vote shares of the far-right leader Marine Le Pen across départements in the 2017 French presidential election. Arguably, both the Leave vote and the support for Le Pen can be described as having a distinct populist flavour. The question is whether both votes are related to similar underlying socio-economic conditions. We find that the factors driving the French vote are indeed similar to those in the United Kingdom. A corresponding model for France using the same variables as for the United Kingdom has explanatory power not far below that for the Brexit referendum. Even an out-of-sample prediction of the French vote using UK estimates performs reasonably well.

This paper and the Brexit vote it studies can be seen not only in an EU context but also related to 'populist' campaigning and voting more broadly. A large literature in the social sciences looks at voting patterns across the political spectrum as a function of demographic, economic and political drivers (see Ferree *et al.*, 2014). The United Kingdom, with its first-past-the-post electoral system for the House of Commons, has typically had clear majorities for either the Conservatives or the Labour Party since the 1920s. This pattern was broken in 2010 with the first coalition government that saw the Conservatives and the Liberal Democrats join forces. Since the 1990s two other major developments have affected the UK party landscape: the rise of the UK Independence Party (UKIP) and the rise of 'nationalist' parties in Scotland, Wales and Northern Ireland. While the latter can be seen as a domestic move driven by a renewed push for devolution (and even independence) for the constituting nations of the United Kingdom, the rise of UKIP is directly related to the EU. Whitaker and Lynch (2011) as well as Clarke *et al.* (2016) study voting patterns for UKIP and document that, not surprisingly, Euroscepticism combined with anti-immigration sentiment is the main driving force of UKIP success. For Western Europe more broadly, Arzheimer (2009) analyses contextual factors explaining far-right voting over the period from 1980 to 2002.

Backlash against globalization is said to have been another important factor in the Leave vote, especially to the extent that it deteriorates economic and social conditions for a subset of voters (see Druckman and Lupia, 2000; Lewis-Beck and Stegmaier, 2000).[5] Colantone and Stanig (2016) provide evidence of import competition from China being related to support for Vote Leave in an arguably causal manner. Their results are consistent with ours in two ways. First, we also find a positive relationship between trade intensity (in our case with other EU countries) and support for Vote Leave. Second, we confirm that areas heavily dependent on manufacturing employment were more likely to vote Leave.

Of course, the UK's Brexit vote should not be equated with support for UKIP or far-right voting more generally. Yet, there are probably some parallels with voting patterns for right-wing parties in other countries and the 'once-in-a-lifetime' opportunity to vote against what many voters see as unaccountable forces ruling them from the outside. In the UK context, Becker and Fetzer (2016) explore the impact of immigration from Eastern Europe on the support for UKIP. Dippel *et al.* (2015) link votes for far-right parties in Germany to trade integration with China and Eastern Europe. For the United States, Autor *et al.* (2016) argue that rising trade integration with China contributes to the polarization of US politics. Burgoon (2012) analyses party opposition and support for trade openness across the EU. Barone *et al.* (2016) find that in Italy, immigration generates a sizeable causal increase in votes for the centre-right coalition that has a political platform less favourable to immigrants.

The UK's EU referendum is of course also related to research on referenda as a form of direct voting. While countries such as Switzerland have ample experience in 'direct voting' (see Funk and Gathmann, 2015), referenda in other countries are rather rare. The United Kingdom traditionally respects the primacy of Parliament over any direct voting. But both the UK's EEC referendum in 1975 and the EU referendum in 2016 were initiated by the House of Commons. Theoretical research has come up with suggestions to improve the efficiency of referenda (Casella and Gelman, 2008). On the empirical side, Matsusaka (1992) asks why some issues are resolved by popular vote and others by elected representatives. Using data on California he finds that 'good government' issues were usually resolved by legislative measures and distributional issues by initiatives. In light of this finding, it makes sense to view the Brexit referendum as one that was at least partially related to distributional issues.

Our paper is organized as follows. Section 2 introduces our empirical approach. Section 3 discusses the underlying data and our main hypotheses. In Section 4, we present our results. Section 5 provides a summary and policy conclusions.

5 Findlay and O'Rourke (2007) argue that globalization was historically difficult to maintain unless domestic institutions developed and adapted accordingly. This often meant a strong role for the state, for instance in the form of educational, training and welfare programmes.

2. EMPIRICAL APPROACH

We take a comprehensive approach to understanding the factors behind the EU referendum result, and we exploit a range of data sources in the empirical analysis. We would like to stress right away that our analysis cannot possibly establish causality. Instead, we try to capture predictive power of various groups of regressors to see which factors explain a larger share of the variation in the Vote Leave share. This approach is quite natural in this setting with a once-in-a-lifetime referendum where we are bound to analyse cross-sectional variation only. If we were to analyse general election results, we could recur to difference-in-difference type estimates in order to control for fixed effects at the local authority level. In our analysis, we do not necessarily expect coefficient signs of each and every coefficient to be stable across all specifications. Instead, it is expected that the signs of some regressors, to the extent that they are highly correlated with each other, may change when 'more fundamental' regressors are added. We will discuss all these issues in more detail when interpreting our results.

We carry out three main exercises: a full model, a best subset selection (BSS) procedure and a within-city analysis. We describe these here in turn. Readers familiar with model selection procedures may want to jump to Section 2.3.

2.1. Full model

The first approach aims at building a 'full' empirical model of the correlation structure between k-dimensional cross-sectional covariates X_c at the local authority area level (380 spatial units across England, Scotland and Wales) and a dependent variable y_c, which is either the share of votes to leave L_c or turnout T_c.[6]

For time-varying observables, the cross-sectional covariates contain their respective baseline levels (mostly from the 2001 census), x_{c0}, as well as their changes, Δx_c, mostly between 2001 and 2011, the two census years. The empirical specification takes the form

$$y_c = \mathbf{x}_c'\beta + \epsilon_c, \tag{1}$$

which we estimate with ordinary least squares (OLS).

2.2. Model selection

In the second approach, we perform a variable selection exercise to identify the most robust predictors of the Vote Leave result. In order to identify robust predictors of the

6 We remind readers that we drop Northern Ireland and Gibraltar. Northern Ireland is dropped because referendum results were only published for the whole region, at a much more aggregated level than for all other parts of the United Kingdom. Gibraltar, a British Overseas Territory, has many missing covariates. Our results are robust to the inclusion of these additional observations.

Vote Leave result, we perform a BSS procedure. BSS is a machine learning method used to perform 'feature selection' in settings where the aim is to reduce dimensionality of a feature space (Guyon and Elisseeff, 2003). The idea of BSS is to estimate all possible regressions including all combinations of control variables and return the statistically optimal model, which minimizes an information criterion.

The fundamental difference between prediction, which generally takes advantage of machine learning methods, and causal inference is as follows. While causal inference focuses on the internal validity of causally estimated reduced-form (or structural) parameters β, prediction and thus machine learning is concerned with the external validity of the estimated fitted values $\hat{y}$. Causal inference seeks to obtain a set of estimated parameters $\hat{\beta}$ that are typically studied in isolation. Thus, they often do not render themselves useful for predictive exercises since the out-of-sample model fit is generally poor. Instead, good model fit typically requires a multitude of regressors, and machine learning can often substantially improve out-of-sample predictive performance (Mullainathan and Spiess, 2017).[7] The underlying estimated parameters that yield good model fit are typically of limited interest per se.[8]

We note that the variables we consider pass a first plausibility test (as they were mentioned during the campaign, e.g.). They cover broad socio-economic characteristics. They are related to the political science literature documenting determinants behind elections (we refer to that literature in Section 1). They do not contain 'nonsensical' variables that could be thought of as generating 'random' and thus meaningless correlations.

The BSS algorithm we employ finds the solution to the following non-convex combinatorial optimization problem:

$$\min_{\beta} \underbrace{\sum_{c=1}^{C} (y_c - \beta_0 - \sum_{j=1}^{p} x_{cj}\beta_j)^2}_{\text{Residual sum of squares}} \text{ subject to } \sum_{j=1}^{p} \mathbf{I}(\beta_j \neq 0) \leq s, \tag{2}$$

where p is the set of regressors of which a subset s is chosen to maximize overall model fit. The result is a sequence of models $\mathcal{M}_1, \ldots, \mathcal{M}_s, .., \mathcal{M}_p$, where the overall optimal model $\mathcal{M}_{s*}$ is chosen by using either cross validation or some degree-of-freedom-adjusted measure of goodness of fit such as the Akaike information criterion (AIC). Throughout, we use the AIC to decide upon the overall optimal model $\mathcal{M}_{s*}$ robustly explaining the variation in the dependent variable.

It is easy to see that this statistically optimal procedure can quickly become infeasible. Suppose there are p potential regressors. BSS proceeds as follows: the first model estimates – using OLS – all $\binom{p}{1} = p$ different models containing a single regressor and

7 See Section 4.5 where we predict out-of-sample the results of the 2017 French presidential election.

8 Some machine learning methods are non-parametric to the extent that the methods do not even produce any model parameters in a classical regression sense.

chooses as optimal the model that results in the largest reduction in the residual sum of squares. The second model estimates all possible $\binom{p}{2}$ models containing exactly two regressors, and so on. In total, $\sum_{k=1}^{p}\binom{p}{k} = 2^p$ models are estimated. With $p = 30$ this amounts to estimating just over 1 billion regressions. The non-feasibility of BSS for large p in high dimensional data has led to machine learning research efforts focusing on developing algorithms that solve an approximation of the BSS optimization problem such as Lasso, Ridge regression or Forward/Backward stepwise selection [see Hastie *et al.* (2009) for an overview].

It is important to highlight that the BSS approach may yield models of different complexity that are non-nested. We present the sequence of 'best' models for each class of models with p predictors and explore how the inclusion of more covariates expands the goodness of fit. One caveat with this approach is that certain variables may be dropped in case they are highly correlated with each other. That is, even if a predictor x_i contains a distinct signal conditional on x_j, it may be dropped from the analysis as the signal contained is not sufficiently strong.

2.3. Within-city analysis

While official results are only published at the level of local authority areas, we also managed to obtain voting data at the ward level across four UK cities (see Section C.1 in Online Appendix C for a description). This allows us to zoom into city wards. It also allows us to address potential worries about ecological fallacy. There is ample variation in the Vote Leave shares within cities. As a matter of fact, the variation within cities is larger than across local authorities.

2.4. (No) difference-in-differences

We considered using the 1975 EU referendum in a difference-in-differences framework. Unfortunately, corresponding data for the 1975 referendum were only published for 68 counting areas across the United Kingdom (see Online Appendix Figure A1 for a map of the Leave vote in the 1975 referendum). More importantly, the 1975 referendum took place in a completely different environment. At the time, the Labour party had pledged to hold a referendum. Margaret Thatcher, the newly elected leader of the Conservatives at the time, campaigned for Remain. Remain won with a smashing 67.2% vote share. Against this backdrop, a difference-in-difference analysis is not possible. Note, however, that we include the 1975 referendum vote shares as a regressor in our analysis and generally find a negative correlation between the 1975 Leave share and the 2016 Leave share. This finding attests to the notion that these referenda took place under very different circumstances [see Butler and Kitzinger (1976) and Crafts (2016) for further background].

3. HYPOTHESES AND DATA

In this section, we discuss prominent hypotheses that have been proposed to explain the EU referendum result and how we try to capture them in our empirical analysis. We briefly discuss the variables employed in the analysis.

The empirical analysis of UK election data is challenging as the data are provided only at the relatively coarse geographic resolution of 380 local authority areas.[9] We start out in Section 3.1 by discussing our main outcome variable, the Vote Leave share in the 2016 referendum, as well as turnout and then turn to the explanatory factors behind the outcome. For these factors, we will look at four broad groups of variables:

1. EU exposure through immigration, trade and structural funds;
2. local public service provision and fiscal consolidation;
3. demography and education;
4. economic structure, wages and unemployment.

We also look at 'random events' on the referendum day such as rainfall and train cancellations. We discuss each group of variables in Sections 3.2–3.6. Table A1 in the Online Appendix provides summary statistics for our variables (not standardized).

Finally, in Section C.1 we also describe data used for an analysis at the level of wards within four UK cities. Wards are areas of finer geographical disaggregation, essentially city quarters, with an average population of about 7,000 (compared with roughly 170,000 residents per local authority area).

Since we are engaged in a prediction exercise and not in a structural estimation of voting behaviour, we are agnostic about whether voting results are better explained by levels of predictor variables, or by changes in those variables over a longer period. Therefore, throughout the analysis whenever available we generally use both levels and changes.[10]

3.1. Voting outcomes

We collect data on turnout and vote shares at the local authority level for the 2016 EU referendum held on 23 June 2016. Vote Leave won 51.9% of votes in the EU referendum, with a standard deviation of 10.4% across UK local authority areas. A total of 46.5 million voters were registered, and 72.2% of these turned out. Thus, 17.4 million voted for Leave and 16.1 million for Remain. These numbers correspond to 37.4% and 34.7% of eligible voters, respectively.

9 Due to missing covariates, we drop Northern Ireland and Gibraltar from the available maximum of 382 areas. A few covariates are also missing for some additional local authority areas, which is why some specifications in our regression tables contain fewer observations.

10 As a robustness check, we use levels and changes separately in Tables A3 and A4 in the Online Appendix.

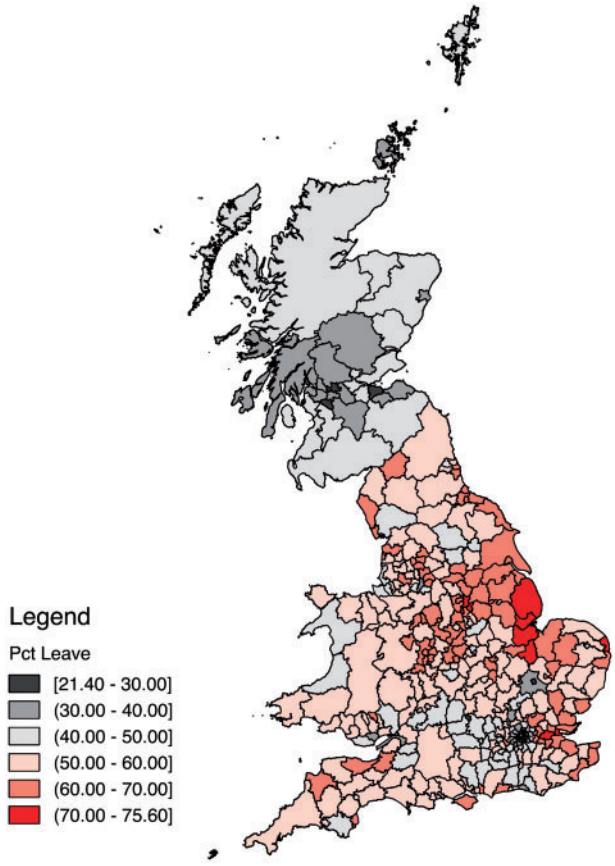


Figure 1. Map of the Leave share (in %) across local authority areas in the 2016 EU referendum.

Figure 1 presents a map of the support for the Leave side across local authority areas, while Figure 2 presents the map pertaining to turnout. One striking observation is that some urban centres seemed to have particularly low turnout. Within London, six local authority areas (the City of Westminster along with the Boroughs of Newham, Camden, Lewisham, Tower Hamlets, Barking and Dagenham) had turnout of less than 65% (out of a total of only 22 local authority areas across the whole of the United Kingdom). Since support for Remain in the EU was strongest in London, low turnout could potentially have affected the overall margin of the result. In Section 4.4, we will discuss speculative scenarios to see how likely differential turnout is in explaining the result.

While our analysis is cross-sectional in nature, it is interesting to note that the 2016 EU referendum result is closely correlated with the UKIP vote share in the 2014 European Parliament elections, as illustrated in Figure A2 in the Online Appendix.[11]

11 Also see Goodwin and Heath (2016).

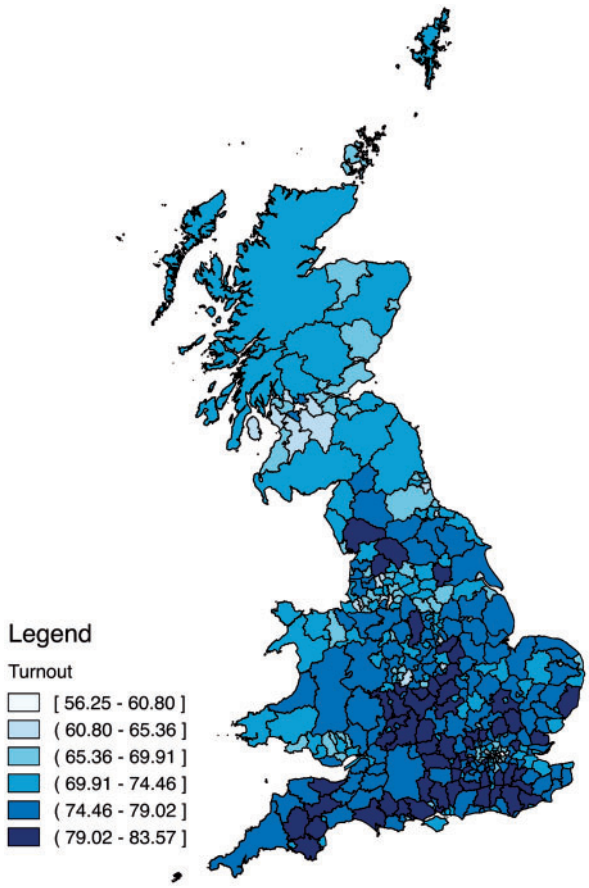


Figure 2. Map of turnout (in %) across local authority areas in the 2016 EU referendum.

The positive relationship is striking. A simple regression line has an intercept of around 25% and a slope close to unity, yielding an R^2 of 75%.[12] While it is beyond the scope of our correlational analysis to uncover the true causal relationships, the tight link suggests that the evolution of UKIP support over time may provide a lens for understanding the causal drivers behind the EU referendum result [see Becker and Fetzer (2016) for an analysis of UKIP vote shares in EP elections in 1999, 2004, 2009 and 2014].

3.2. EU exposure: immigration, trade and EU transfers

In a referendum on EU membership, the most natural predictors for the decision to remain in or leave the EU are variables that capture the UK's exposure to the EU.

12 In the working paper version of this paper, Becker *et al.* (2016), we also used UKIP vote shares in the regression analysis.

Depending on the costs and benefits from EU membership that different parts of the country perceive, measures of immigration, trade and receipt of EU structural funds are likely to matter for the Vote Leave share.

3.2.1. Immigration. We first consider immigration, a central topic throughout the Leave campaign. In the wake of the Eastern enlargement of the EU in 2004, the United Kingdom, Ireland and Sweden were the only countries not to impose transitional controls on migrants from new member states. The United Kingdom only put in place immigration controls when Bulgaria and Romania joined the EU in 2007, but those elapsed by 2014. Given that UK wages are a multiple of those in accession countries, many Eastern European workers moved to the United Kingdom, and immigration has been at the forefront of the public debate ever since, especially in the tabloid press. While net immigration from the EU to the United Kingdom was only 15,000 in 2003, in the year before Eastern enlargement, it jumped to 87,000 in 2004. It fell slightly in the aftermath of the global financial crisis when pound sterling depreciated, only to rise strongly again to an all-time peak of 184,000 in 2015.[13] Nevertheless, it comes as a surprise to many political observers that the *net* migrant stock with other EU countries is substantially *lower* in the United Kingdom than in Germany, Spain and France, not least because the United Kingdom has a fairly high emigration rate to the EU compared with these countries (Vargas-Silva, 2012).

In fact, immigration has ranked as a top priority for UK voters over the last decade, together with the economy and the National Health Service (NHS). A key pillar of the Leave campaign was to promise control of immigration by restricting the free movement of labour from other EU countries. However, throughout that period net immigration from non-EU countries always exceeded EU net immigration typically by a substantial margin, especially prior to 2013 (see Wadsworth *et al.*, 2016).[14]

To capture the trends in immigration, we link data from the 2001 and 2011 censuses on levels as well as growth rates in the local resident shares by three origin groups (EU 15 countries, the 12 EU accession countries that joined the EU in 2004 and 2007, and non-EU migration).[15]

3.2.2. Trade. The 'take back control' theme of the Leave campaign also extended to the free movement of goods and services. Many voters perceived international trade not as

13 Figures are from the https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration Office for National Statistics.

14 In a string of recent immigration-related referenda in Switzerland, the rural regions that had comparatively little immigration tended to vote most strongly against it, see http://www.migrationpolicy.org/article/switzerlands-non-eu-immigrants-their-integration-and-swiss-attitudes here. Likewise, EU migrants are heavily concentrated in London where the Remain vote share was particularly high.

15 The migration growth rate is defined as the change in the number of migrants between 2001 and 2011 relative to the local resident population in 2001. Our migration data are by country of birth, not by citizenship. That means first-generation immigrants from earlier migration waves (e.g., from Commonwealth nations in the 1950/60s) are captured if still alive in 2011.

an opportunity to sell to foreign markets but rather as unwelcome competition threatening their jobs and livelihoods. To address the role played by 'globalization' and 'foreign competition' in the context of international trade, we match data on EU trade integration of individual UK regions to local authority areas. Specifically, we measure trade integration as the share of value added in a UK region that can be attributed to consumption and investment demand in the rest of the EU. These data are available by 37 NUTS2 regions in the United Kingdom for the year 2010. There is considerable variation across UK regions. The highest degree of trade integration can be found in East Yorkshire and Northern Lincolnshire, Cumbria, Leicestershire, Rutland and Northamptonshire (over 14%), and the lowest in Inner London, North Eastern Scotland, Eastern Scotland and the Highlands and Islands (around 4%).[16] We stress that for the purposes of interpreting our regression results in Section 4, it is important to keep in mind that due to the higher aggregation at the NUTS2 level, we have in principle less variation in our trade integration measure.[17]

3.2.3. EU transfers. Lastly, a further central topic of the referendum campaign was the size of British EU budgetary contributions. The Leave campaign quoted a figure suggesting that every week £350 million were sent to Brussels as the UK's contribution to the EU budget. This figure was widely criticized as misleading since a significant share of the funds were returned to the United Kingdom (the net contribution was closer to £120 million per week).[18] While the gross payment towards the EU budget is not attributable to voting areas, we can track funding received from the EU. Data on EU funding are available by 133 regions in the United Kingdom. Those are essentially NUTS3 regions but were aggregated in a few cases because of past changes to boundaries of NUTS3 regions. We map them onto the local authority areas. On the one hand, EU funding has been found to be generally beneficial to regional growth (Becker *et al.*, 2010, 2012, 2013). But on the other hand, EU funding may be perceived by voters as a handout and a symbol of foreign dependence (Davies, 2016).

3.3. Public service provision and fiscal consolidation

The referendum also presented an opportunity for those 'left behind' to express their anger, more generally speaking. The Vote Leave promise of 'taking back control' lent itself to an interpretation beyond control of borders and was seen as invitation to take back

16 We source the data on value-added shares from Los *et al.* (2017). It combines the contributions of all major sectors to regional GDP (services, manufacturing, construction and primary industries including agriculture, mining and energy supply). Los *et al.* (2017) find a *positive* correlation between EU trade integration and the share of voters intending to vote Leave.

17 See Arnorsson and Zoega (2016) for an analysis of the Brexit vote at the level of those NUTS2 regions. However, these authors do not use any trade-related covariates.

18 The £350 million number is even incorrect as a gross figure since it does not account for the UK rebate.

control of their own lives and express anger over a ruling class that has not addressed reduction congestion of public services, whether or not related to immigration.

3.3.1. Fiscal cuts. In the wake of the global financial crisis, the coalition government brought in wide-ranging austerity measures to reduce government spending and the fiscal deficit. At the level of local authorities, spending per person fell by 23.4% in real terms from 2009/10 until 2014/15. But the extent of cuts varied dramatically across local authorities, ranging from 46.3% to 6.2% with the sharpest cuts typically in the poorest areas (Innes and Tetlow, 2015). It is important to note that the variation of cuts across local authorities is driven by the unequal share of the population that receives different kinds of benefits, hence cuts are generally larger in more deprived areas. Given this, it is not surprising that in regressions where we control for demographic characteristics that capture 'need', the fiscal cuts coefficient changes substantially, reflecting the more fundamental nature of the underlying demographics that are themselves predictors of those cuts. While some spending budgets such as the NHS were ring-fenced and therefore experienced small or no cuts, other areas such as social services and housing benefits faced drastic spending reductions. At the same time, a growing population and immigration further increased pressure on public services.

We obtain data compiled by the Financial Times capturing the geographic heterogeneity of budget cuts across all UK local authority areas. These variables capture various spending cuts affecting housing benefits, non-dependant deductions, disability living allowance, incapacity benefits, child benefits and tax credits. The measures are expressed in terms of the financial loss per working adult in pounds sterling per year over the period from 2010 to 2015. The overall financial loss per working adult varies between £914 in Blackpool and £177 in the City of London. Most fiscal cuts were applied across the board affecting individual claimants across the country fairly homogeneously. This implies that the geographic variation in the size of the fiscal cuts captures the underlying baseline degree of demand for benefits: the places with highest demand for benefits were naturally more affected.[19] In other words, fiscal cuts largely reflected (and reinforced) weak fundamentals (see also Beatty and Fothergill, 2016).

3.3.2. NHS service delivery. The Leave campaign made frequent reference to the pressure on public services in general and the NHS in particular, mainly holding immigration responsible although in fact, immigrants from the EU were net contributors and thus subsidized public spending and helped to reduce the fiscal deficit (Wadsworth *et al.*, 2016).

As a measure of NHS service delivery we capture the fraction of suspected cancer patients who are being treated within 62 days from being first seen by a doctor. This is a key NHS health target metric for which we obtained data for the fourth quarter of

19 The data are available http://ig.ft.com/austerity-map/ here and explained in more detail http://ig.ft.com/austerity-audit/methodology.html here.

2015/16 across England, Scotland and Wales.[20] We match the local authority areas to 230 clinical commission groups under the oversight of the NHS Commissioning Board Authority. The fraction of treated patients varies from around 60% to 90%.[21]

3.3.3. Pressure on the housing market. Immigration is often made responsible for pressures on the housing market, which is suffering from a structural deficit of newly built properties especially across the growing urban centres in the South. We therefore complement the fiscal consolidation and NHS waiting time variables with data from the 2001 and 2011 censuses on the shares of the population owning a house (outright or mortgaged), or living in council-provided rental housing.

3.3.4. Commuting. In addition, we use 2011 census data to control for the share of working age residents that commute to Inner London for work. Commuting is supposed to capture two things: first, it can be seen as 'lack of job opportunities' at place of residence. Second, it measure the luxury enjoyed by those with well-paid jobs in London who reside in posh suburbs. The effect of this variable on the Vote Leave share is ex ante unclear.

3.3.5. Public sector jobs. Furthermore, we consider the public employment share as measured by the Business Register and Employment Survey. This is another important measure of local service provision and jobs under threat in the light of austerity policies.

3.4. Demography, education and life satisfaction

It has been argued that older voters were more prone to Vote Leave, while younger voters overwhelmingly supported Remain. Also, less educated voters are those who might find it harder to grasp the opportunities from globalization in the form of EU membership and at the same time suffer most from the challenges posed by globalization. Voters dissatisfied with their lives and or regions with large disparities in life satisfaction may have been more prone to Vote Leave. We try to capture those factors as follows.

3.4.1. Age structure. To reflect characteristics of the local population, we rely on data from the 2001 and 2011 censuses on the share of the local population by age brackets.[22]

20 The NHS publishes waiting times for a host of potential treatments, but the data for suspected cancer patients were by far the most complete and constitute a treatment that is of particular urgency where prolonged waiting times can have life-threatening consequences.

21 We compute the average within a local authority area. If no clinical commission group sits in a local authority area, we take the value of the nearest one. Patients might choose not to receive treatment (unobservable to us), thereby affecting the overall fraction of treated patients.

22 Those brackets are under the age of 30 years, between 30 and 44 years, between 45 and 59 years, 60 years and older. We ultimately use the share variable for the age group 60 years and older as our reference group. As discussed already, BSS – while powerful – is also prone to a curse of dimensionality problem so that we cannot use an endless number of covariates.

3.4.2. Education. We capture the education of the local population by the shares of people with various qualification levels.[23] Figure A3 in the Online Appendix provides a map of the population shares with no qualifications in the year 2001. We note that, to the extent that education and the age structure of the population are more fundamental factors, it will not be surprising to find that they pick up some of the variation in other 'intermediate' predictor variables of the Vote Leave share: as argued above, fiscal cuts were largely fiscal cuts to benefits enjoyed by older and less educated parts of the population. Also, migration from Eastern Europe was largely into less educated areas (see Becker and Fetzer, 2016), so again we expect variation in education to affect the coefficients on migration variables when all of those variables are pooled in the same regression.

3.4.3. Life satisfaction. We obtained so-called 'headline estimates' of personal well-being from the Annual Population Survey (APS) provided by the Office of National Statistics, available at the level of local authorities, for the year stretching from April 2015 to March 2016. We use both the mean life satisfaction as well as the coefficient of variation over the four categories Low, Medium, High and Very High.

3.5. Economic structure, wages and unemployment

A typical narrative is that the Leave campaign resonated particularly well with voters in areas that had experienced prolonged economic decline, especially in the manufacturing sector. Those at the lower end of the wage distribution might have been more prone to competition from Eastern European migrants, so wages are also a potentially important predictor.

3.5.1. Sector structure. To capture the economic structure across local authority areas we collect data on the employment shares in retail, manufacturing, construction and finance in 2001 and 2011. We use both the employment shares across those sectors in 2001 as well as the changes in those shares between 2001 and 2011 as predictor variables.

3.5.2. Wages. We add information on wages and earnings obtained from the Annual Survey of Hours and Earnings. Specifically, we focus on levels for the year 2005 and

23 There are in principle five brackets: no qualifications, level 1 (up to four GCSEs or equivalent), level 2 (five or more GCSEs or equivalent), level 3 (two or more A levels or equivalent) and level 4 + (undergraduate degree, professional qualification or equivalent). We ultimately use share variables for the lowest and highest qualification levels, the remainder being the reference group.

changes in median wages between 2005 and 2015.[24] Similarly, we include data from the APS/Labour Force Survey, in particular the unemployment rate, the self-employment rate and overall participation rate of the working age population.

3.6. Campaigning and events on the referendum day

Apart from the four broad groups of predictor variables listed so far, events on the day of the poll may also be important in explaining turnout and voting patterns. Heavy rain in London and the Southeast of England led to the cancellation of trains during the evening rush hour, and a number of commuters did not reach the voting booths in time before their 10 pm closure. In line with earlier research (see Madestam *et al.*, 2013; Meier *et al.*, 2016), this weather pattern may potentially influence turnout and the voting result in affected areas. We pair daily rainfall measurements from the CHIRPS precipitation data set, available at a 0.05 degree resolution, with the share of residents in a local authority area who commute to London. We investigate whether significant rainfall had an effect on turnout and the Vote Leave result across local authority areas that host a large share of London commuters.[25]

In addition, we also study the role of the tabloid press. We construct a measure covering the extent to which the Daily Mail, the Sun and the Daily Express are read by residents in these areas. For lack of detailed geographic circulation data, we rely on the British Election Study (BES) data for 2001, 2005, 2010 and 2015. All these surveys contain a question whether an individual reads a daily newspaper and if so, which one it is. We match respondents (who live in wards of sampled constituencies) to the local authority area and compute an average of the number of respondents over all these BES surveys who report reading the Daily Mail, the Sun and the Daily Express.[26] These are naturally noisy proxies and they are only available for around 185 local authority areas, which is why we treat this analysis as a separate exercise.

4. RESULTS

In Section 3, we discussed our variables in different groups. To get a first indication of how these groups are related to the 2016 EU referendum result, in Section 4.1 we first regress the vote shares separately on the variables of each group. Our aim is two-fold. First, discussing groups of variables separately allows us to concentrate on the relative importance of variables within a thematic group as predictors of the Vote Leave result.

24 Bell and Machin (2016) report a negative relationship between median wages and the Vote Leave share.

25 The CHIRPS data are available http://chg.geog.ucsb.edu/data/chirps/ here.

26 We only include local authority areas with at least ten respondents across these four surveys. Restricting the set to only include local authority areas with at least 30 respondents yields very similar results.

In Online Appendix B, we also perform speculative back-of-the-envelope calculations to see by how much important predictor variables would have had to be different in order to overturn the referendum result. Second, looking at the R^2 for groups of variables informs us about the predictive power of thematic groups relative to each other. After this, in Section 4.2, we pool the groups of variables and perform the BSS procedure more generally. Finally, in Section 4.3, we highlight the role played by the interaction of key predictor variables. This allows us to answer questions such as whether fiscal cuts affected the referendum result more in regions with weaker fundamentals.

4.1. Predicting the Brexit vote by variable group

All of the four tables pertaining to results for the four groups of predictor variables (Tables 1–4) follow the same logic: the first column shows the one variable that has the best predictive power among all variables in the variable group. The subsequent columns show the different best subsets for regressions with two regressors (Column 2), three regressors (Column 3), etc. The last column reports the full set of regressors.

It is important to remember that the best subset of $k-1$ predictors is not necessarily nested in the best subset of k predictors. Table 2 is a case in point where the regressor in Column 1 does not appear in Column 2. For this reason, in Tables 1–4 there is no 'triangular' structure for the columns displaying the different best subsets. Note that we standardize all right-hand side variables to mean 0 and a standard deviation of 1 to ease comparability of coefficient estimates. The left-hand side variable is the percentage of the Leave vote, i.e., it varies between 0 and 100.

4.1.1. Group 1: EU exposure (immigration, trade and structural funds). In Table 1, we correlate the Vote Leave share with measures of immigration, EU trade dependence, EU subsidies (Structural Funds) and the 1975 referendum Leave share. The variation from the initial EU 15 migrant resident share in Column 1 alone generates an R^2 of 29.6%. Adding the measure of EU trade dependence in Column 2 increases the R^2 further. These two regressors together have the largest explanatory power of any two variables in this first group of predictors, jointly explaining 42.8% of the variation in the referendum result. The subsequent columns add only marginally to the R^2. Overall, the full set of regressors explains 48.3% of the variation in the Vote Leave share. Using the AIC as our degree-of-freedom-adjusted measure of goodness of fit, Column 6 turns out to provide the best trade-off between parsimony and overall explanatory power. This column is marked by an 'X' in the row 'Best Subset'. All subsequent tables follow the same logic.

We use migrant resident shares in levels for the year 2001 and their growth between 2001 and 2011 for three subgroups: migrants from the 12 EU accession countries that joined in 2004 and 2007, from the initial EU 15 countries and from non-EU countries. It turns out that migrant shares in *levels* are negatively correlated with the Brexit vote as

Table 1. Predictors of Brexit vote: EU exposure (immigration, trade and structural funds)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Initial EU accession migrant resident share (2001)				−1.197	−1.753***	−1.651**	−1.428**	−1.267	−1.271
				(0.767)	(0.657)	(0.645)	(0.708)	(0.870)	(0.871)
EU accession migrant growth (2001–11)					1.138**	1.376**	1.085*	1.276**	1.303*
					(0.522)	(0.533)	(0.554)	(0.632)	(0.663)
Initial EU 15 migrant resident share (2001)	−5.665***	−4.739***	−5.504***	−4.692***	−4.632***	−3.941***	−3.825***	−3.757**	−3.771***
	(0.893)	(0.854)	(1.104)	(1.361)	(1.397)	(1.518)	(1.470)	(1.475)	(1.453)
EU 15 migrant growth (2001–11)						−1.165	−1.120	−0.921	−0.914
						(0.771)	(0.753)	(0.841)	(0.827)
Initial migrants from elsewhere resident share (2001)								−0.570	−0.504
								(0.972)	(1.223)
Migrants from elsewhere growth (2001–11)									−0.102
									(0.859)
Total economy EU dependence (2010)		3.896***	2.586***	2.466***	2.536***	2.395***	2.659***	2.622***	2.616***
		(0.407)	(0.495)	(0.465)	(0.457)	(0.449)	(0.487)	(0.492)	(0.494)
EU structural funds per capita (2013)							0.556	0.525	0.522
							(0.571)	(0.575)	(0.576)
1975 referendum Leave share			−2.401***	−2.356***	−2.259***	−2.121***	−2.046***	−2.038***	−2.040***
			(0.585)	(0.586)	(0.579)	(0.592)	(0.675)	(0.678)	(0.677)
Best subset						X			
Observations	380	380	380	380	380	380	369	369	369
R^2	0.296	0.428	0.464	0.471	0.48	0.485	0.483	0.483	0.483

Notes: Table reports results from OLS regressions. The dependent variable is the share of the Leave vote in a local authority area in England, Scotland and Wales. Empirical models selected using BSS on the set of predictors using the AIC information criterion. Best subset marked by 'X'. Robust standard errors are presented in parentheses, asterisks indicate ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 2. Predictors of Brexit vote: public service provision and fiscal consolidation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of residents commuting to London (2011)	−4.767***			−2.608***	−2.990***	−2.695***	−2.708***	−2.701***
	(0.353)			(0.566)	(0.538)	(0.549)	(0.545)	(0.569)
Owned (outright + mortgage) share (2001)		7.385***	7.267***	5.378***	4.818***	6.120***	6.129***	6.128***
		(0.482)	(0.490)	(0.676)	(0.648)	(0.863)	(0.866)	(0.861)
Owned (outright + mortgage) share growth (2001–11)								0.023
								(0.511)
Council rented share (2001)						1.609***	1.771**	1.762**
						(0.609)	(0.745)	(0.718)
Council rented share growth (2001–11)							0.275	0.280
							(0.613)	(0.625)
Total fiscal cuts (2010–15)		5.370***	5.556***	5.056***	5.802***	5.619***	5.629***	5.637***
		(0.450)	(0.440)	(0.466)	(0.499)	(0.488)	(0.487)	(0.501)
Share of suspected cancer patient treated within 62 days (2015)			−2.186***	−2.654***	−2.433***	−2.398***	−2.377***	−2.381***
			(0.584)	(0.663)	(0.527)	(0.510)	(0.514)	(0.527)
Public employment share (2009)					−2.166***	−2.278***	−2.260***	−2.262***
					(0.590)	(0.583)	(0.588)	(0.579)
Best subset						X		
Observations	376	379	378	375	375	375	375	375
R^2	0.215	0.431	0.475	0.503	0.535	0.544	0.545	0.545

Notes: Table reports results from OLS regressions. The dependent variable is the share of the Leave vote in a local authority area in England, Scotland and Wales. Empirical models selected using BSS on the set of predictors using the AIC information criterion. Best subset marked by 'X'. Robust standard errors are presented in parentheses, asterisks indicate ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 3. Predictors of Brexit vote: demography and education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of resident population no qualifications (2001)				4.939***	7.263***	6.467***	6.445***	6.519***
				(0.745)	(0.898)	(0.844)	(0.834)	(0.902)
Share of resident population no qualifications growth (2001–11)			2.697***	4.215***	5.443***	4.900***	4.938***	4.965***
			(0.436)	(0.562)	(0.588)	(0.568)	(0.560)	(0.586)
Share of resident population qualification 4 + (2001)	−8.208***	−8.159***	−10.103***	−6.540***	−5.763***	−6.149***	−6.030***	−6.024***
	(0.434)	(0.399)	(0.418)	(0.785)	(0.821)	(0.703)	(0.684)	(0.688)
Share of resident population qualification 4+ growth (2001–11)					2.375***	2.049***	1.956***	1.950***
					(0.465)	(0.451)	(0.455)	(0.455)
Population 60 years and older (2001)							0.456*	0.412
							(0.254)	(0.273)
Population 60 years and older growth (2001–11)				2.815***	2.622***	2.186***	2.171***	2.117***
				(0.296)	(0.291)	(0.277)	(0.272)	(0.281)
Mean life satisfaction APS well-being data (2015)								0.135
								(0.379)
CV life satisfaction APS well-being data (2015)		2.650***	2.195***			1.369***	1.300***	1.308***
		(0.293)	(0.273)			(0.236)	(0.237)	(0.239)
Best subset							X	
Observations	380	378	378	380	380	378	378	378
R^2	0.621	0.687	0.722	0.743	0.776	0.795	0.796	0.796

Notes: Table reports results from OLS regressions. The dependent variable is the share of the Leave vote in a local authority area in England, Scotland and Wales. Empirical models selected using BSS on the set of predictors using the AIC information criterion. Best subset marked by 'X'. Robust standard errors are presented in parentheses, asterisks indicate ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 4. Predictors of Brexit vote: economic structure, wages and unemployment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Retail employment share (2001)	7.019***	5.514***	4.302***	4.254***	4.097***	4.304***	4.147***	4.182***	4.045***	3.759***	3.803***	3.814***	3.821***	3.724***	3.721***
	(0.418)	(0.403)	(0.430)	(0.434)	(0.398)	(0.374)	(0.377)	(0.371)	(0.382)	(0.401)	(0.402)	(0.403)	(0.434)	(0.450)	(0.456)
Retail employment share change (2001–11)										−0.594	−0.601	−0.593	−0.591	−0.601	−0.605
										(0.429)	(0.429)	(0.428)	(0.443)	(0.444)	(0.447)
Manufacturing employment share (2001)		3.621***	3.688***	3.516***	5.405***	5.498***	5.632***	6.051***	5.901***	5.955***	5.916***	5.901***	5.909***	5.797***	5.786***
		(0.356)	(0.302)	(0.317)	(0.509)	(0.500)	(0.498)	(0.591)	(0.625)	(0.611)	(0.613)	(0.615)	(0.650)	(0.659)	(0.687)
Manufacturing employment share change (2001–11)					2.237***	2.478***	2.537***	2.734***	2.591***	2.317***	2.362***	2.354***	2.363***	2.319***	2.316***
					(0.546)	(0.547)	(0.540)	(0.553)	(0.537)	(0.600)	(0.603)	(0.606)	(0.661)	(0.665)	(0.665)
Construction employment share (2001)			3.220***	3.203***	3.014***	3.042***	3.304***	3.338***	3.226***	3.254***	3.328***	3.314***	3.317***	3.390***	3.391***
			(0.426)	(0.417)	(0.418)	(0.411)	(0.426)	(0.422)	(0.450)	(0.441)	(0.469)	(0.481)	(0.495)	(0.506)	(0.510)
Construction employment share change (2001–11)				1.326***	1.529***	1.643***	1.473***	1.412***	1.414***	1.336***	1.380***	1.376***	1.375***	1.419***	1.425***
				(0.384)	(0.380)	(0.395)	(0.402)	(0.407)	(0.428)	(0.413)	(0.409)	(0.408)	(0.411)	(0.411)	(0.425)
Finance employment share (2001)								0.586	0.961**	0.945**	1.063**	1.075**	1.068**	0.988**	0.986**
								(0.429)	(0.423)	(0.419)	(0.437)	(0.440)	(0.451)	(0.459)	(0.463)
Finance employment share change (2001–11)											0.325	0.349	0.355	0.342	0.342
											(0.430)	(0.428)	(0.443)	(0.441)	(0.441)
Median hourly pay (2005)													0.059	−0.228	−0.244
													(1.063)	(1.225)	(1.243)

(*continued*)

Table 4. Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Median hourly pay							−0.843**	−0.786**	−1.108***	−1.071***	−1.103***	−1.123***	−1.112***	−1.090**	−1.092**
change (2005–15)							(0.338)	(0.331)	(0.369)	(0.371)	(0.374)	(0.366)	(0.397)	(0.481)	(0.486)
Interquartile pay range									−0.861	−0.932*	−1.094**	−1.136**	−1.175	−0.912	−0.897
(2005)									(0.535)	(0.551)	(0.551)	(0.574)	(0.866)	(1.016)	(1.038)
Interquartile pay range														0.081	0.081
growth (2005–15)														(0.448)	(0.449)
Unemployment rate						0.827***	0.873***	0.875***	0.736**	0.692**	0.639*	0.688**	0.688**	0.707**	0.703*
(2015)						(0.303)	(0.304)	(0.302)	(0.321)	(0.325)	(0.327)	(0.344)	(0.345)	(0.352)	(0.361)
Self-employment rate															−0.027
(2015)															(0.412)
Participation rate												0.132	0.128	0.210	0.215
(2015)												(0.364)	(0.379)	(0.386)	(0.383)
Best subset										X					
Observations	380	380	380	380	380	377	377	377	369	369	369	369	369	366	366
R^2	0.454	0.554	0.637	0.653	0.667	0.674	0.68	0.682	0.693	0.695	0.696	0.696	0.696	0.695	0.695

Notes: Table reports results from OLS regressions. The dependent variable is the share of the Leave vote in a local authority area in England, Scotland and Wales. Empirical models selected using BSS on the set of predictors using the AIC information criterion. Best subset marked by 'X'. Robust standard errors are presented in parentheses, asterisks indicate ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

those immigrants predominantly moved to urban areas that subsequently voted for Remain in 2016. The striking observation is that in terms of migrant share *growth*, only migration from the mainly Eastern European EU accession countries *positively* correlates with the Vote Leave share. The well-established literature studying the economic implications of migration on labour market outcomes supports the notion that there are distributional consequences of low-skilled migration putting pressure on wages for low-skilled natives (see e.g., Borjas, 2003; Cortes, 2008; Borjas and Monras, 2016). Migration from Eastern Europe, predominantly of low-skilled workers, affected areas with a lower-skilled resident population.[27] As we will see below, low skills correlate with a larger Vote Leave share.

In terms of the point estimates, their interpretation is simplified by the fact that all regressors are standardized to have mean 0 and a standard deviation of 1. For instance, in the best subset specification displayed in Column 6, a one-standard deviation higher initial EU 15 migrant share is associated with a 3.941 percentage-point lower Vote Leave share.

In Online Appendix B we explore, in a speculative way, what may have happened to the EU referendum vote under alternative scenarios in which migration to the United Kingdom would have been different. We find that since the vote shares do not appear very sensitive to migration, only a large reversal of the EU accession immigration experience would have swayed the vote. We stress, however, that such speculative scenarios must be taken with a large grain of salt, not least since various regressors on the right-hand side are correlated and a causal interpretation is generally not possible.

The EU trade dependence of local authority areas is also positively correlated with the Vote Leave share. The reason is that areas with a heavy concentration of manufacturing (such as the North East of England) tend to disproportionately import from and export to EU countries, and those areas were likely to vote Leave. This finding has been highlighted in the public discussion before: those areas most dependent on trade integration with the EU were more likely to vote Leave (see Los *et al.*, 2017). Interestingly, shortly after the referendum when Nissan threatened to stop further investment in Sunderland (one of the areas with a large Vote Leave share), pressure mounted on Westminster to do 'something' to keep Nissan on board.

EU Structural Funds per capita over the EU Programming period 2007–13 have no predictive power. Some have argued that EU subsidies in the form of EU Structural Funds would 'buy votes.' Davies (2016) argues that EU funding may be perceived by voters as a handout and a symbol of foreign dependence. As a consequence, regions receiving more money may loathe the EU more. Interestingly, Cornwall, the area receiving the largest amount of EU Structural Funds per capita, voted Leave but immediately

27 Becker and Fetzer (2016) estimate the causal effect of immigration from Eastern Europe on the UKIP vote share in European Parliament elections, which, as we saw above, strongly correlates with the Vote Leave share.

after the referendum (on 24 June 2016) pleaded with the UK government to continue payments after EU money runs out. Our results indicate that, on balance, EU Structural Funds do not predict the Vote Leave share.

Finally, we include matched vote shares from the 1975 EU referendum as an additional regressor. There is a strong negative association between voting Leave in 2016 and 1975, suggesting different underlying attitudes and considerations across voting areas (see Section 2.4).

4.1.2. Group 2: Public service provision and fiscal consolidation. In Table 2, we observe that the share of residents in a local authority area who commute to London is a strong predictor for voting Remain.[28] This might be explained by the fact that those commuting into London are relatively high-skilled who have a larger tendency to vote Remain. On the other hand, house ownership is strongly correlated with the Vote Leave share. This correlation may not be surprising as house ownership is highest among the older section of the population. The share of the population in rented council housing, a measure of those potentially under increased pressure from migration of largely low-skilled Eastern European migrants, also has a strong positive correlation with the Vote Leave share.

Another important predictor in this group of variables is the extent of total fiscal cuts. Local authorities experiencing more fiscal cuts are more likely to vote in favour of leaving the EU. Importantly, fiscal cuts were implemented as de-facto proportionate reductions in grants across all local authorities (Innes and Tetlow, 2015). This setup implies that reliance on central government grants is a proxy variable for deprivation, with the poorest local authorities being more likely to be hit by the cuts. This makes it impossible in the cross-section (and challenging in a panel) to distinguish the effects of poor fundamentals from the effects of fiscal cuts. With this caveat on the interpretation in mind, our results suggest that local authorities experiencing more fiscal cuts were more likely to vote in favour of leaving the EU. Given the nexus between fiscal cuts and local deprivation, we think that this pattern largely reflects pre-existing deprivation. In Online Appendix B, we provide speculative scenarios for fiscal cuts.

In a similar manner, pressure on the public health system matters. In regions where the share of suspected cancer patients waiting for treatment for less than 62 days is larger, the Vote Leave share is lower. By symmetry, where waiting times are longer, Vote Leave gains. Finally, areas with a larger share of the workforce in public employment, a measure of (a) availability of public services and (b) public jobs, the Vote Leave share is lower. In summary, results indicate that provision of public services and the severity of fiscal cuts mattered for the referendum result. Overall, variables capturing public service provision and fiscal consolidation explain slightly more than 50% of the variation in the Vote Leave share.

28 Note that people commute to London from as far as Manchester, 200 miles from London and a 2-h train ride from city centre to city centre.

4.1.3. Group 3: demography and education. In Table 3, we explore whether demography and education variables predict the referendum result. As predictors, we use both the baseline levels in 2001 and the growth between 2001 and 2011 of the share of the population that has no qualifications or a high qualification, respectively. The middle qualification range is the reference group. The results indicate that a larger baseline share of the population with no qualifications is associated with a larger Vote Leave share. A stronger increase in that share between 2001 and 2011 is further associated with a higher Vote Leave share. In contrast, the share of the population that has a high qualification is associated with a lower Vote Leave share. But somewhat surprisingly, faster growth of the share with a high qualification is associated with a larger Vote Leave share. We cannot exclude that this partially captures a generally faster increase in the population, which in turn might be associated with pressure on housing and public services.

In terms of age brackets, we use the share of the population aged 60 years and older, which makes those younger than 60 years the reference group.[29] Both a higher baseline share of older people as well as a larger increase in their share between 2001 and 2011 predict a larger Vote Leave share. This is consistent with polls in the run-up to the referendum indicating a clear age gradient in the Vote Leave share, with younger voters intending to vote Remain and older voters intending to vote Leave.[30]

We also add life satisfaction scores from the well-being questions in the APS. The mean score is insignificant. However, the coefficient of variation is positively related to the Vote Leave share. This finding suggests that a higher relative dispersion of well-being across voting areas, which can be interpreted as a measure of life satisfaction inequality, has positive predictive power for the Vote Leave share.

Overall, it is striking that the demography and education group of variables has the largest predictive power of any of the groups, with an R^2 of close to 80% and strongly significant associations in most cases between our regressors and the Vote Leave share.

4.1.4. Group 4: Economic structure, wages and unemployment. In Table 4, we concentrate on variables characterizing the sectoral structure of voting areas, both in terms of levels in the baseline year 2001 and in terms of their changes from 2001 to 2011. We single out employment in retail, manufacturing, construction and finance, and subsume all other sectors in the residual reference category. This reference category is of course quite heterogeneous, containing sectors such as agriculture, the public sector and various service sectors. This being said, a higher share of employment in the baseline year in any of the four sectors highlighted in Table 4 is associated with a larger Vote Leave share compared with the reference category.

29 Note that in principle, we could use more finely grained age brackets. But in the long specifications in Section 4.2, this would run into dimensionality issues for the machine learning algorithm, as explained above.

30 In Online Appendix B, we provide speculative scenarios for qualifications and age.

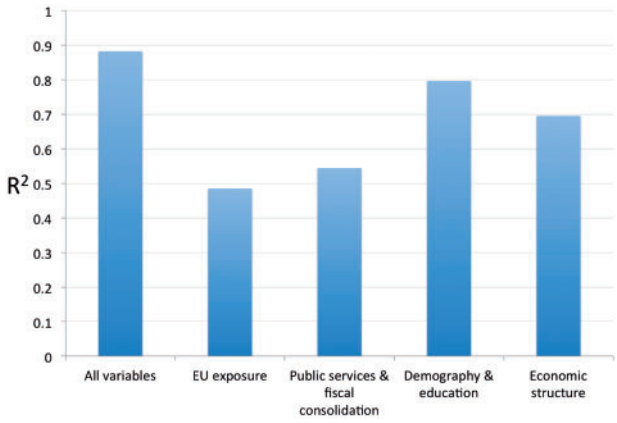


Figure 3. Goodness of fit (measured as R^2) in separate regressions explaining the Leave vote shares at the local authority area level using only regressors from the respective group of variables.

As for the change in employment between 2001 and 2011, a stronger increase in manufacturing, construction and finance employment is associated with a higher Vote Leave share. The growth of retail employment is not significantly associated with the Vote Leave share.

We also include median hourly pay as well as the interquartile pay range as a measure of inequality, again both in terms of levels and their changes (with 2005 and 2015 as the relevant years). A higher median hourly pay in the year 2005 is not significantly related to the vote. However, a stronger increase in that variable is associated with a lower Vote Leave share, consistent with the narrative that those 'left behind' were more likely to vote Leave. We mostly do not find a significant relationship for the interquartile pay range, if anything a negative relationship in levels.

Finally, we add the unemployment rate, the self-employment rate and the general labour participation rate in the year prior to the referendum. A larger unemployment rate is associated with a larger Vote Leave share, but the self-employment and participation rates have no predictive power for the Vote Leave share. Overall, variables in this group explain around 69% of the variation in the Vote Leave share.[31]

4.1.5. Summary of analysis of four groups of predictor variables. Overall, each of Tables 1–4 yields an R^2 of at least 48% with a full set of regressors. The strongest explanatory power lies with demography and education variables in Table 3. Figure 3 gives a visual overview of the goodness of fit across Tables 1–4, while as a comparison the first bar represents the explanatory power of the regression underlying Column 2 in Table 5.

31 In Online Appendix B, we provide speculative scenarios for manufacturing employment and unemployment.

Table 5. Predictors of Brexit vote: blocked variable selection approach

	Combined		Different best subsets			
	(1)	(2)	(3)	(4)	(5)	(6)
Initial EU accession migrant resident share (2001)	−1.678*** (0.530)	−1.722*** (0.597)	−1.651** (0.645)			
EU accession migrant growth (2001–11)		−0.501 (0.425)	1.376** (0.533)			
Initial EU 15 migrant resident share (2001)	2.698*** (0.503)	2.820*** (0.554)	−3.941*** (1.518)			
EU 15 migrant growth (2001–11)		−0.532 (0.562)	−1.165 (0.771)			
Total economy EU dependence (2010)	1.100*** (0.256)	0.947*** (0.282)	2.395*** (0.449)			
1975 referendum Leave share	−0.916*** (0.315)	−0.855** (0.346)	−2.121*** (0.592)			
Share of residents commuting to London (2011)	0.908** (0.426)	0.930* (0.549)		−2.695*** (0.549)		
Owned (outright + mortgage) share (2001)	3.273*** (0.572)	2.950*** (0.583)		6.120*** (0.863)		
Council rented share (2001)	0.650* (0.381)	0.608 (0.411)		1.609*** (0.609)		
Total fiscal cuts (2010–15)	−1.463*** (0.455)	−1.084** (0.544)		5.619*** (0.488)		
Share of suspected cancer patient treated within 62 days (2015)	−0.380 (0.282)	−0.411 (0.279)		−2.398*** (0.510)		
Public employment share (2009)		−0.234 (0.275)		−2.278*** (0.583)		
Share of resident population no qualifications (2001)	6.024*** (0.648)	6.740*** (0.904)			6.445*** (0.834)	
Share of resident population no qualifications growth (2001–11)	2.206*** (0.435)	2.715*** (0.542)			4.938*** (0.560)	
Share of resident population qualification 4 + (2001)	−5.897*** (0.793)	−4.716*** (1.039)			−6.030*** (0.684)	
Share of resident population qualification 4+ growth (2001–11)		0.351 (0.392)			1.956*** (0.455)	
Population 60 years and older (2001)		−0.537 (0.346)			0.456* (0.254)	
Population 60 years and older growth (2001–11)		0.075 (0.339)			2.171*** (0.272)	
CV life satisfaction APS well-being data (2015)		0.146 (0.253)			1.300*** (0.237)	

(*continued*)

Table 5. Continued

	Combined		Different best subsets			
	(1)	(2)	(3)	(4)	(5)	(6)
Retail employment share (2001)	0.689**	0.839**				3.759***
	(0.317)	(0.391)				(0.401)
Retail employment share change (2001–11)	−0.375	−0.177				−0.594
	(0.256)	(0.301)				(0.429)
Manufacturing employment share (2001)		0.802				5.955***
		(0.543)				(0.611)
Manufacturing employment share change (2001–11)		0.866				2.317***
		(0.547)				(0.600)
Construction employment share (2001)		0.473				3.254***
		(0.417)				(0.441)
Construction employment share change (2001–11)	0.664**	0.604*				1.336***
	(0.304)	(0.325)				(0.413)
Finance employment share (2001)	−0.787**	−0.573				0.945**
	(0.326)	(0.362)				(0.419)
Median hourly pay change (2005–15)	−0.455*	−0.514**				−1.071***
	(0.235)	(0.241)				(0.371)
Interquartile pay range (2005)	0.931**	0.502				−0.932*
	(0.434)	(0.448)				(0.551)
Unemployment rate (2015)	0.475*	0.472*				0.692**
	(0.267)	(0.264)				(0.325)
Observations	366	366	380	375	378	369
R^2	0.879	0.882	0.485	0.544	0.796	0.695

Notes: Table reports results from OLS regressions. The dependent variable is the share of the Leave vote in a local authority area in England, Scotland and Wales. Empirical models selected using BSS on the set of predictors using the AIC information criterion. Column 1 shows the best subset across all four groups of variables analysed in Tables 1–4. Column 2 is the full specification based on the best subsets determined in Tables 1–4. For comparison, Columns 3–6 re-display the optimal specifications from Tables 1 to 4. Robust standard errors are presented in parentheses, asterisks indicate ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

The analysis of variables by group mainly served the purpose of considering different aspects of the referendum result in more detail and to see how well different groups of variables perform relative to each other. But of course, it makes sense to allow all groups of variables to 'compete' against each other in a single setup. This is what we turn to in Section 4.2.

4.2. BSS results

In Table 5, we use the BSS procedure for variables across *all* groups. Column 1 displays the best subset of variables when all the 'best' variables from the four separate groups of regressors are combined in one joint 'horse race.' The regressors include two migration variables, EU trade dependence, the 1975 referendum vote share, fiscal cuts, various qualification variables, median pay and the unemployment rate, among others. Overall, we obtain an R^2 of almost 88% with 19 variables.

Column 2 displays a full specification including all variables without performing another round of BSS that yields essentially the same R^2, despite the fact that the model of Column 1 is a restricted version of the model in Column 2. As a comparison, Columns 3–6 re-display estimates using only the best subsets uncovered in each of the four variable groups from the previous tables. We stress that as in previous tables, Table 5 just reports conditional correlations with no causal identification.

We need to point out one caveat when it comes to the interpretation of Column 2 of Table 5. While the point estimates, coefficient signs and statistical significance of variables *within variable groups* are internally consistent when we add successive regressors (using the same procedure underlying Tables 1–4), some coefficient signs and statistical significance patterns are different in the combined model of Column 2 compared with Columns 3–6. This is not surprising per se. The differences are attributable to the tight correlation between regressors *across variable groups*. For example, in Column 2 the coefficient on total fiscal cuts is negative in contrast to the positive coefficient in Column 4.

In particular, the demographic variables are tightly correlated with other key variables of interest. For example, the correlation between the share of individuals with no qualifications and the fiscal cuts measure is 65%. Similarly, the growth in the share of individuals with low qualifications may be partly driven by low-skilled migrant growth (its correlation with EU accession migrant growth is 48%). Hence, it is not surprising that when we remove the qualification measures from the analysis, the coefficient patterns across fiscal cuts and EU accession migration growth remain stable (see Table A2 in the Online Appendix in contrast to Table 5).

For completeness and as a robustness check, we also perform a BSS exercise focusing on variables in levels (see Table A3 in the Online Appendix) and variables in changes (see Table A4 in the Online Appendix). In our baseline Table 5 we have two sets of regressors. First, we have a common core of variables that are in levels only. Second, we have a set of variables for which both changes as well as baseline levels are available (mostly qualification variables and employment shares). Table A3 performs BSS on the first set and the subset of the second set of variables that are levels only. Online Appendix Table A4 performs BSS on the first set and the subset of the second set of variables that are changes only. Given the smaller range of variables to choose from in each table, it is not surprising that overall explanatory power in terms of R^2 is lower in principle. But it still turns out roughly the same as in the case of Online Appendix Table A3. For the most part, certainly in Online Appendix Table A3, the variables show similar patterns of magnitude and significance as in Table 5.

To understand not only the predictive but rather the causal drivers of the Brexit vote, it would seem important to analyse data in panel form. We highlight that political support for the UKIP party in previous European Parliament elections, due to its strong predictive power for the Leave vote in the 2016 referendum, might be the appropriate outcome measure to better understand the causal mechanisms by which other characteristics affect the 2016 referendum result. Becker and Fetzer (2016) provide a first attempt along those lines, studying the effect of migration from Eastern Europe on UKIP vote

shares over time. It seems an important future research agenda to use plausible identification strategies and possibly micro-level data on individual voters to explain voting patterns in response to changes in socio-economic fundamentals.

Finally, we also consider the voting results separately for Scotland only. As there are only 32 voting areas in Scotland, we face lower statistical power and hence a larger number of insignificant coefficients. Nevertheless, we find broadly similar regression results in terms of signs and relative magnitudes compared with those in Tables 1–5 for the entire sample. In particular, we find similar roles for higher qualification and median pay (associated with a lower Vote Leave share) and higher manufacturing employment (associated with a higher Vote Leave share).[32] Therefore, while the intercept of support for Vote Leave is clearly lower in Scotland, we do not have evidence to suggest that the coefficient patterns (i.e., slopes) for Scotland behave very differently from those for the entire sample.

Section C.2 in Online Appendix C documents that similar socio-economic forces also seem to be associated with the Vote Leave result when we explore within-city variation. This suggests that the underlying associations do not just mask a divide between urban and rural areas.

4.3. Interaction terms

While we have so far concentrated on a comprehensive approach to predicting the Vote Leave share, we also want to highlight whether salient factors reinforced each other. In the debate before and after the referendum, increased migration and fiscal cuts were highlighted as two salient developments over the years preceding the vote. Arguably, migration and fiscal cuts might have had a stronger influence on the Vote Leave share when hitting areas with different pre-existing conditions. In other words, we would like to see whether the interaction of local area characteristics influenced the degree to which migration and fiscal cuts influenced the Vote Leave share. Of course, we cannot carry out such an exercise for all of the variables entering our analysis so far, so we take an eclectic approach. We look at the flow (i.e., growth) of new migrants from Eastern European EU accession countries, the flow of new migrants from 'old' EU member countries and the flow of new migrants from outside the EU, as well as our measure of total fiscal cuts as 'flow' variables in separate regression analyses.

Each of these flow variables are interacted with one of the following 'stock' variables: the share of the population with no qualifications; the sectoral share of manufacturing; the sectoral share of finance, all three measured in 2001; the median hourly pay in 2005. The results are striking and highlight some important aspects. The main effects of the stock variables that characterize 'pre-existing conditions' in the first row of Table 6

32 We do not include those results here but they are available upon request.

Table 6. Pairwise interactions

	Flow: EU accession migration				Flow: EU 15 member country migration				Flow: Migration from non-EU				Flow: Total fiscal cuts			
Stock:	No Qualification	Manufacturing	Finance	Wage	No Qualification	Manufacturing	Finance	Wage	No Qualification	Manufacturing	Finance	Wage	No Qualification	Manufacturing	Finance	Wage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Stock	7.048***	5.745***	−3.223***	−4.786***	6.238***	4.832***	−1.424**	−3.415***	6.805***	5.431***	−2.284***	−4.052***	10.499***	5.909***	−3.461***	−5.687***
	(0.330)	(0.482)	(0.636)	(0.435)	(0.338)	(0.483)	(0.658)	(0.506)	(0.345)	(0.449)	(0.707)	(0.479)	(0.525)	(0.479)	(0.703)	(0.670)
Flow	−1.034***	0.229	−0.752	−0.808**	−3.924***	−1.858	−5.941***	−4.987***	−1.935***	−0.882	−2.487***	−1.791***	−4.790***	0.619	1.601***	−0.292
	(0.372)	(0.684)	(0.604)	(0.366)	(0.544)	(1.442)	(0.664)	(0.750)	(0.432)	(0.686)	(0.641)	(0.560)	(0.639)	(0.497)	(0.489)	(0.532)
Interaction	1.679***	1.411*	−2.349***	−3.048***	−0.435	1.277	0.620**	0.572**	0.187	0.960*	−0.418	−1.052**	−0.210	−0.368	−1.965**	−1.544***
	(0.277)	(0.757)	(0.661)	(0.382)	(0.312)	(0.915)	(0.272)	(0.259)	(0.329)	(0.564)	(0.502)	(0.459)	(0.373)	(0.467)	(0.759)	(0.490)
Observations	380	380	380	380	380	380	380	380	380	380	380	380	379	379	379	379
R^2	0.527	0.346	0.153	0.31	0.591	0.433	0.292	0.357	0.525	0.348	0.163	0.276	0.61	0.326	0.174	0.269

Notes: Table reports results from OLS regressions. The dependent variable is the share of the Leave vote at the local authority area level. The table presents the results for a range of interaction effects, interacting pre-determined 'stock' variables measured in 2001 (for the share of households with no qualifications, the share of employment in manufacturing and finance) and in 2005 for the median wage with a range of 'flow' variables capturing migration growth between 2001 and 2011 and the extent of fiscal cuts. Robust standard errors are presented in parentheses, asterisks indicate ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

are consistent across all four different 'flow' variables: the share of the population with no qualifications and the share of those in manufacturing are both associated with a larger Vote Leave share, whereas the share of those in finance and a higher median hourly pay tend to be associated with a lower Vote Leave share. Turning to the main effect of the flow variables, migration from any origin region is, if anything, negatively associated with the Vote Leave share. The main effect for fiscal cuts differs across stock variables.

Most importantly, the interaction terms, which are the main focus here, show a striking pattern. A larger flow of migrants from Eastern Europe reaching a local authority area with a larger share of unqualified people or a larger share of manufacturing workers is associated with a larger Vote Leave share, whereas the opposite is true when a large flow of migrants from Eastern Europe reaches an area with a large share of those working in finance, or an area with higher median hourly pay. In other words, initial conditions matter.

The pattern is less clear for migration flows from 'old' EU 15 countries and from non-EU countries. Here, point estimates on the interaction terms are generally smaller and often insignificant. This suggests that migration from Eastern Europe, which was distinct in nature by consisting of more lower-skilled migrants, had a different effect.

Interestingly, the interaction terms of fiscal cuts with the share of unqualified or manufacturing workers are insignificant. At first sight this non-finding may be seen as surprising since anecdotally, the significant welfare reforms and cuts were politically contentious, and the Leave campaign implicitly suggested that the UK's contributions to the EU budget should be used to fund the UK's welfare system instead. Our interpretation for this non-result is as follows: most of the cuts that were implemented by David Cameron's government were not explicitly discriminatory but rather applied homogeneously across the United Kingdom. Since the demand for benefits is strongly associated with weak fundamentals such as a workforce with low qualifications, this implies that the incidence of cuts in per capita terms is strongly correlated with these weak fundamentals. In fact, the correlation between the share of the population with no qualifications and the total fiscal cuts measure is 65%, indicating that there is little independent variation that may be captured by an interaction effect. However, looking at the interaction between fiscal cuts and the finance share of the workforce and the median hourly pay variable, we find that larger fiscal cuts fostered a larger Vote Leave share in areas with a smaller finance sector and lower wages.

4.3.1. The role of media exposure. We described the fact that data on media exposure is available for only less than half of local authorities. Still, many readers will be keen on understanding the role that media exposure played for the UK referendum result. Table 7 concentrates on understanding the link between education and media exposure. Arguably, less-educated voters may be more susceptible to 'negative press' in the form of anti-EU propaganda by the likes of the Daily Mail. Column 3 of Table 7 shows that turnout is neither significantly associated with the main effect of media exposure

Table 7. Tabloid press penetration, education and the EU referendum

	Turnout			Pct Leave		
	(1)	(2)	(3)	(4)	(5)	(6)
Daily Mail/Sun/ Express penetration	0.886**	0.416	0.411	5.649***	1.871***	1.745***
	(0.395)	(0.263)	(0.269)	(0.814)	(0.465)	(0.447)
Share of resident population no qualifications (2001)		0.410	0.379		2.981**	2.270
		(1.213)	(1.242)		(1.450)	(1.468)
Share of resident population qualification 1 (2001)		1.151*	1.141*		−1.694**	−1.906**
		(0.662)	(0.670)		(0.789)	(0.792)
Share of resident population qualification 2 (2001)		3.948***	3.949***		3.725***	3.742***
		(0.477)	(0.477)		(0.592)	(0.606)
Share of resident population qualification 4 + (2001)		3.379**	3.329**		−5.283***	−6.432***
		(1.543)	(1.588)		(1.891)	(1.824)
Daily Mail/Sun/ Express penetration × Share of resident population no qualifications (2001)			0.061			1.402***
			(0.309)			(0.401)
Observations	185	185	185	185	185	185
R^2	0.0281	0.594	0.594	0.255	0.803	0.817

Notes: Table reports results from OLS regressions. The dependent variable in Columns 1–3 is turnout as the share of the registered electorate in a local authority area that cast their votes, while in Columns 4–6 it is the Vote Leave share. Newspaper penetration was constructed from the BES data for 2001, 2005, 2009 and 2015. Robust standard errors are presented in parentheses, asterisks indicate ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

nor with the interaction term with education in Column 3. However, Column 6 shows that Daily Mail/Sun/Express penetration has a positive association with the Vote Leave share, both as a main effect and even more so among the least educated.[33]

4.4. Turnout as dependent variable

While our main analysis is concerned with the Vote Leave results, it is also instructive to look at turnout as an alternative outcome.[34] Table A6 in the Online Appendix presents those results. For the sake of brevity, we just briefly highlight a few results. Columns 1

33 For completeness, we perform a best sample selection exercise including our media variable but on a smaller sample due to the missing observations (see Online Appendix Table A5). Media exposure shows up as significant with a positive sign. Otherwise, results are fairly similar to the baseline findings in Table 5.

34 Given the regional nature of our analysis, we cannot say much about the motivation of individual voters to turn out. Empirical evidence using individual-level data suggests that social norms, peer pressure and monitoring play a key role in voter participation (see e.g., Gerber *et al.*, 2008; DellaVigna *et al.*, 2017). For theoretical considerations on turnout and quora, see Herrera and Mattozzi (2010) and Levine and Mattozzi (2017).

Table 8. Did bad weather affect the referendum result?

	Turnout		Pct Leave	
	Rainfall amount (1)	Rainfall top decile (2)	Rainfall amount (3)	Rainfall top decile (4)
Inner London commuters	1.310	−0.052	−6.306***	−5.380***
	(0.834)	(0.413)	(1.266)	(0.475)
Rainfall on 23 June	1.025***	2.330**	1.584***	2.560
	(0.309)	(0.979)	(0.588)	(2.090)
Inner London commuters	−1.879***	−2.162***	0.408	0.304
× rainfall on 23 June	(0.455)	(0.552)	(0.718)	(0.803)
Observations	372	372	372	372
R^2	0.137	0.07	0.228	0.219

Notes: Table reports results from OLS regressions. The dependent variable in Columns 1 and 2 is turnout as the share of the registered electorate in a local authority area that cast its vote, while in Columns 3 and 4 it is the Vote Leave share. Rainfall data are drawn from the CHIRPS rainfall data product. Robust standard errors are presented in parentheses, asterisks indicate ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

and 2 indicate that areas which experienced strong immigration growth from EU accession countries had higher turnout. Areas that had a higher support for Leave in the 1975 referendum (which tend to be areas that were more in favour of Remain in the 2016 referendum, see Table 1) had lower turnout. On balance, the results therefore suggest that turnout was lower in those areas with a higher potential in favour of Remain.

Column 4 shows that areas with more deprivation, as measured by stronger fiscal cuts, had lower turnout. Similarly, Column 6 shows that areas with higher unemployment also had lower turnout. In contrast, areas with an older population and higher wages had higher turnout (see Columns 5 and 6, respectively).

4.4.1. Rainfall. Moreover, we study the extent to which bad weather across commuting zones south of London affected the EU referendum result. Rainfall led to train cancellations and may have had an influence by disproportionately reducing turnout of voters who commute into London and may have been more likely to harbour pro-EU preferences, given the strong overall support for Remain in London. The results are presented in Table 8. Results for turnout as the dependent variable are shown in Columns 1 and 2, and results for the Leave share as the dependent variable are shown in Columns 3 and 4. The findings suggest that the combination of rainfall and commuting into London is indeed associated with significantly lower turnout but not with significantly different vote shares. However, given our turnout scenarios for turnout below, any reasonable change in turnout behaviour across London commuting areas would not have been sufficient to overturn the referendum result.

4.4.2. Speculative scenarios for turnout. Finally, we consider turnout scenarios (not based on regression results). According to detailed polling conducted after the

referendum, turnout for the bracket of youngest voters aged 18–24 years was 64%.[35] This compares to turnout for the same age group of less than 50% on average in UK general elections since 2000, and to an average turnout in the referendum across all age groups of 72.2%. Turnout for voters aged 25–39 years was 65% and thus also higher than in previous general elections but by a smaller margin. On the other end of the age spectrum, voters aged 65 years and above had a turnout of 90%. Support for Leave steadily increased with age, rising from just 27% for 18–24 year-olds to 60% for voters aged 65 years and above.[36]

Could the referendum have ended up with a victory for Remain if more young people had turned out? We first focus on the required increase in turnout by voters aged 18–24 years only. We use population shares by age from the Office for National Statistics from 2015 in combination with the above data on turnout and Leave support by age group. The age group of 18–24 year-olds makes up around 11.3% of the voting population. Holding fixed population shares, Vote Leave shares and the turnout of all other age groups, we calculate that turnout among 18–24 year-olds would have had to be approximately 120% instead of 64% to overturn the referendum result. Clearly, this would not have been feasible.

How about a broader group of Remain voters? According to Ashcroft (2016) the bracket of voters aged 35–44 years still voted Remain by 52%, while the next bracket of voters aged 45–54 years voted majority Leave. Could the referendum have gone the other way with a higher turnout among all voters up to the age of 44 years? We calculate the across-the-board increase in turnout in that larger age bracket which would have been required to overturn the result. This increase would have been 32 percentage points. That is, instead of the turnout of roughly 65% among voters up to the age of 44 years, a turnout of 97% would have been required. Of course, this is unrealistic.

Overall, we therefore conclude that higher turnout among the youngest section of the voting population, or even among all age groups that voted majority Remain, would not have overturned the referendum outcome.

4.5. Out-of-sample prediction for the 2017 French presidential election

Since our results establish correlative patterns in the data and do not allow for a causal interpretation, the question arises as to whether our results can be useful for forecasting. In particular, can our results predict other election outcomes in an out-of-sample manner?

To address this issue we consider the 2017 French presidential election. It is interesting to compare it against the UK Brexit referendum since arguably both votes featured

35 See Helm (2016) for the turnout figures by age group in the referendum and Burn-Murdoch (2016a) for turnout in previous general elections.

36 See Ashcroft (2016) for vote shares by age group.

strong 'populist' movements. In the Brexit case, the role of UKIP was fundamental in making the referendum possible in the first place. UKIP also played a key rule during the referendum campaign. In the French case, the Front National led by Marine Le Pen was a key contender.

As the dependent variable corresponding to the Leave share in the Brexit referendum, we consider the vote share for the Front National candidate Marine Le Pen, both in the first as well as in the second round of the French presidential election in April/May 2017. We examine the Le Pen vote share at the level of 95 French départements.[37] As to the right-hand side variables, we take the variables selected in the best subset in Column 1 of Table 5 as the baseline specification. Our aim is to construct the corresponding French variables as closely as possible.

Due to data limitations we have to adjust some variables as follows. The French data do not allow us to distinguish between EU 15 migrants and migrants from the 12 EU Eastern accession countries. We therefore construct an EU 27 variable that captures both groups. We also construct the corresponding UK variable. Moreover, our French wage change variable is based on average wages across French départements (median hourly pay in the United Kingdom). Instead of the interquartile pay range, we use the slightly more compressed 70th/30th percentile range in France (but we keep the interquartile variable name). A further caveat is that these data are only available at the NUTS1 region level as opposed to the département level. For the employment shares, we have to rely on the French working-age population as a denominator rather than the resident population. There are minor discrepancies in terms of the years. For instance, the French house ownership data are for 2013 (2001 for the United Kingdom). The French qualifications growth data are for 1999 and 2013 (2001 and 2011 for the United Kingdom) and are only provided across three groups (as opposed to four distinct qualification groups in the UK case). The French migrant share refers to 2008 (2001 for the United Kingdom), which is the year France fully opened its borders to migrants from the 2004 EU accession countries. In addition, we drop the 1975 referendum variable since there never was such a referendum on EU membership in France. We also drop the variables on the share of residents commuting to London, the council rented share, cancer patients and fiscal cuts since we were unable to find corresponding French data. Finally, we standardize our regressors as before, but we also standardize and thus demean the dependent variables so that they are measured in directly comparable units.

As our first step, we run regressions with these updated variables for both the UK and French samples. We report the results in Online Appendix Table A7. Columns 1, 3 and 5 include the full set of regressors. Columns 2, 4 and 6 show the BSS. For the UK specifications in Columns 1 and 2 we obtain an R^2 of almost 85%, while for the French specifications in Columns 3–6 we obtain an R^2 around 65%. There is no major difference

37 We do not have sufficient data for the overseas départements and Corsica, which leaves us with 95 départements in mainland France.

between the first and second rounds of the French presidential election. Overall, the chosen variables therefore pass the plausibility test of explaining the French data fairly well. Given that the variables are initially chosen based on the UK data underlying Table 5, it is to be expected that we achieve higher explanatory power for the UK specifications.[38]

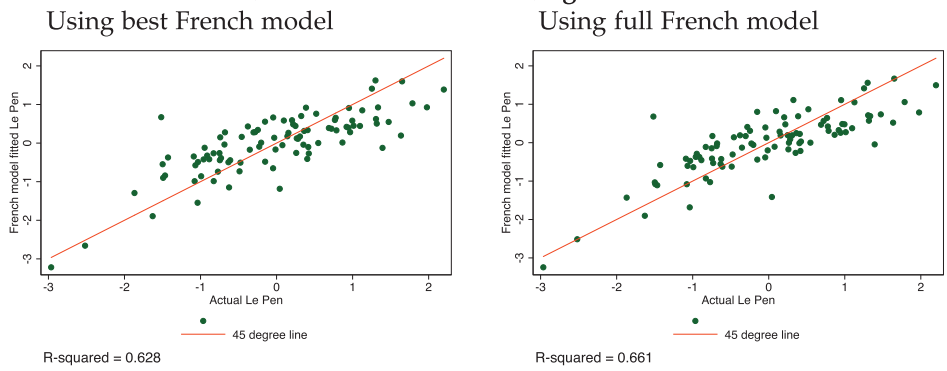
Many coefficients are very similar (and significant) across the UK and French specifications, in particular EU trade dependence and no qualifications. But in France unemployment played a more important role. Looking at the individual coefficients, we see that some of the strongest (absolute) magnitudes are found for the qualification variables, both in the United Kingdom and the French context. They tend to be substantially larger, for instance, compared with EU trade dependence. This pattern underlines the relative importance of education.

We now turn to our main objective, which is the out-of-sample prediction. We illustrate our results in Figure 4. Panel A is based on Columns 3 and 4 of Online Appendix Table A7. It compares the fitted values of the Le Pen vote estimated off French data against the actual French data without imposing UK coefficient values yet (we use the first-round Le Pen shares for the purpose of Figure 4). The graph on the left-hand side is based on the best model (Column 4) while the right-hand side is based on the full model (Column 3), but the fit is roughly the same. Panel A serves as a benchmark in the sense that if we use the French data to predict the French outcome without any coefficient constraints, we obtain an R^2 of 63% in a regression of the fitted values on the actual values. In Panel B we show predicted values based on the UK coefficients applied to the French data, plotted against the actual French data. This is thus a constrained version of Panel A. We therefore obtain a lower R^2 of 33%, which is roughly halved. That is to say, using the model that is constrained to use the coefficients estimated off the UK data, we are still able to explain roughly 50% of the variation that the best empirical model could achieve based on the set of covariates we have available for France. This is the main result of the out-of-sample prediction exercise. Finally, Panel C compares the fitted values from panel A against the predicted values from Panel B. The correlation here is higher than in Panel B with an R^2 of 50%. This tells us that the two predictions are relatively more closely related, meaning that individual observations tend to deviate from the true observations in similar ways.

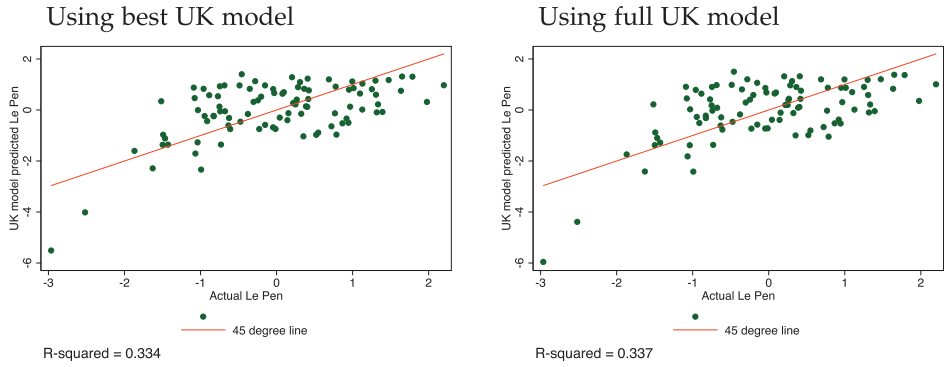
Overall, we conclude that the model we estimate for the Leave shares in the UK Brexit referendum is not purely idiosyncratic. It seems that similar factors are at work for the Le Pen vote shares in France. Naturally, the French model performs best when we estimate France-specific coefficients. When we constrain the underlying coefficients to the UK values, the explanatory power is approximately halved but clearly, there are systematic similarities between the UK and French votes.

38 Note that since we have standardized the dependent variables in Online Appendix Table A7, the coefficients are not directly comparable to those in Table 5.

Panel A: **Fitted values (based on French model) against actual values**



Panel B: **Predicted values (based on UK model) against actual values**



Panel C: **Predicted values against fitted values**

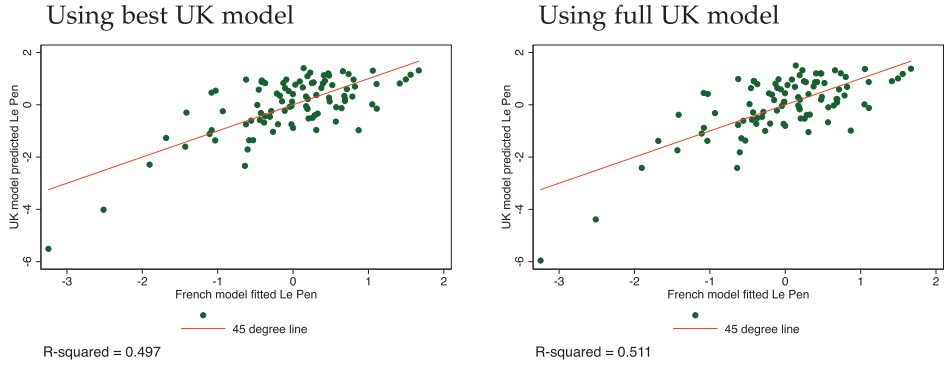


Figure 4. Results from the Le Pen vote share prediction (based on vote shares in the first round of the 2017 French presidential election).

Notes: Panel A compares the fitted values of the Le Pen vote estimated off French data against the actual French data. Panel B compares the predicted values based on the UK coefficients applied to the French data against the actual French data. Panel C compares the fitted values from panel A against the predicted values from panel B.

5. SUMMARY AND POLICY CONCLUSIONS

5.1. Policy conclusions

In terms of policy conclusions, we argue that the voting outcome was driven by long-standing fundamental determinants, most importantly those that make it harder to deal with the challenges of economic and social change. They include a population that is less educated, older and confronted with below-average public services. A complex picture arises about the challenges of adapting to social and economic change – challenges that differ across local authority areas. These spatial disparities might be reinforced by people self-selecting into local areas that better fit their outlook on life, for instance socially liberal professionals concentrating in London. This self-selection might explain the perceived increase in political polarization between 'cosmopolitan' and 'provincial' areas (Jennings and Stoker, 2016).

As economic change is often driven by global trend and developments, it is in our view an important avenue for future research to better understand the relationship and interplay between domestic and international politics, in particular in the context of the supranational institutions such as the EU. Rodrik (2016) highlights the tension between democracy and ever more globalization if national sovereignty is supposed to be maintained. Müller (2016) argues that a lack of genuine political choice can foment populism and the rise of authoritarian parties who claim that they alone can speak on behalf of the 'real people' and their true interests.

5.2. Polls, betting markets and the Westminster bubble

One key question remains. If the voting outcome seems relatively clear with hindsight, why did it come as such a surprise during the referendum night? Some Remain supporters highlighted the possibility of a Leave majority early on, for example the prominent Labour politician Andy Burnham from the Northwest of England as early as March 2016.[39] But the majority of journalists and politicians seem to have been caught off guard, including staff running the Remain campaign.[40] There is some evidence that when it comes to sensitive issues, individuals are more likely to reveal their true opinions if polls are double-blind. We therefore expect that the Brexit referendum (and also the US election in November 2016) will have important implications for polling methods and survey methodology.

39 Burnham warned of 'too much Hampstead and not enough Hull.' See http://www.express.co.uk/news/politics/652503/Burnham-admits-Britains-WILL-vote-to-leave here and http://www.theguardian.com/politics/2016/jun/10/andy-burnham-warns-remain-is-failing-to-reach-labour-heartland here.

40 See Peter Mandelson's account of the Remain campaign http://on.ft.com/297qF6M here and also http://www.politico.eu/article/how-david-cameron-lost-brexit-eu-referendum-prime-minister-campaign-remain-boris-craig-oliver-jim-messina-obama/ here.

Similarly, throughout the campaign betting markets predicted the wrong outcome, typically showing a strong majority for Remain. As most money in total was wagered on Remain (although a large number of small bets were placed on Leave) and as betting markets balance the books, it is perhaps not surprising that betting markets did not get it right. The confidence in a Remain victory was also at odds with the polls, which suggested a much tighter race. In fact, analysing 121 opinion polls in the run-up to the referendum, Clarke *et al.* (2016) suggest that 'Leave was almost certainly ahead of Remain over the entire last month of the campaign – and possibly throughout 2016.'

It is clear that a substantial subset of politicians and the media were genuinely surprised by the referendum result. This speaks to the polarization between metropolitan and other areas. We find it plausible that the 'Westminster bubble' may play a part in understanding the voting outcome, in combination with inaccurate polling. The underrepresentation of anti-EU parties in the British parliament is likely a crucial contributing factor to the lack of attention in the political process paid to struggling areas, especially in England and Wales. As a result of the first-past-the-post voting system, UKIP currently has no Member of Parliament in the House of Commons out of over 600, despite the fact that UKIP came first in the most recent European Parliament elections. UKIP representatives are therefore hardly in positions of political responsibility and thus largely escape media scrutiny. It may therefore be appropriate to consider ways of improving the diversity of views represented in British politics.

5.3. Could other countries follow the British and leave the EU?

Leaving the EU amounts to a major constitutional change for the United Kingdom. Given how much British politics has struggled with political decisions that are relatively minor in comparison, for instance the expansion of Heathrow Airport or the HS2 high-speed rail network, it is astonishing how such a far-reaching constitutional matter appears to have been decided by a referendum with no more than a simple majority and without an initial parliamentary debate on the same question (Kinsman, 2016).[41] These circumstances may be unique to the United Kingdom. France, for instance, requires constitutional revisions to be passed by both houses of parliament with subsequent approval through a referendum, or by a three-fifths parliamentary majority.

In any case, the United Kingdom has always had a more ambiguous relationship with the EU, having been denied entry twice through French vetoes (see Online Appendix A). Margaret Thatcher negotiated the UK budget rebate in 1984. The United Kingdom opted out of the Euro and the Schengen Agreement and has looser arrangements regarding the Charter of Fundamental Rights and areas of freedom, security and justice.

41 The UK Parliament only voted on the EU Referendum Act in 2015 but at the time did not debate the substance matter of EU membership.

Nevertheless, could Brexit be followed by Frexit for Grexit? Our analysis shows how the United Kingdom is characterized by stark differences across local areas in terms of the vote outcome and underlying factors such as economic structure, education and immigration growth. Facing declining incomes and the challenge of adapting to a rapidly changing environment in terms of structural change and immigration, it may not be surprising that voters in some areas seized the opportunity to lash out at the established political order (O'Rourke, 2016). Similar trends of decline and structural change in parts of the economy can be observed in other EU countries. Indeed, analysing the vote shares for the far-right leader Marine Le Pen in the 2017 French presidential election, we find similar driving forces at work. While specific political circumstances may always be unique to each country, we do not see any a priori reasons to believe that it would be impossible for a similar backlash to happen elsewhere in Europe.

Whatever Brexit option the UK pursues, Britain's EU referendum can be seen as a protest from those feeling left behind and dissatisfied with the state of politics. Politicians in other European countries would be wise to heed the call.

Discussion

Vasso Ioannidou

Lancaster University

In the summer of the 2016, the British electorate stunned the world when it voted to leave the EU. In an effort to understand who voted for Brexit, this paper offers a systematic analysis of the socio-economic characteristics that correlate with the outcome of the 2016 referendum and in particular with the cross-sectional variation in the Vote Leave shares and turnout shares across the United Kingdom's 380 local authority areas.

The analysis explores the role of nearly 30 indications, capturing four broad factors: (i) an area's exposure to the EU and other countries (e.g., immigration and trade), (ii) the quality of public services provision and exposure to fiscal consolidation, (iii) an area's demographic and human capital characteristics (e.g., age, education, and life satisfaction), and (iv) an area's underlying economic structure (e.g., unemployment rate, wages, and inequality). These four groups of factors are evaluated in isolation (i.e., one group at a time) and against each other (i.e., all together).

The results indicate that a simple empirical model can explain a (stunningly) large amount of the variation in the Vote Leave share across local authorities. Interestingly, and contrary to the political debate in the run-up to the election, the authors find that relatively little variation in the Vote Leave share is explained by an area's exposure to the EU or other countries (immigration and trade) or by the quality of public services and fiscal consolidation. Demographic and education characteristics are instead found

to matter decisively more. Two variables, age and education, can explain as much as 77% of the cross-sectional variation in the Vote Leave share. One variable – the share of population with no qualifications – can explain as much as 62% of the variation in the Vote Leave share! No other variable, among the nearly 30 variables considered, is found to have such a strong explanatory power. What explains this incredibly high explanatory power of education?

It is hard say. Reduced-form associations such as those uncovered in the paper do not necessarily capture causal relationships. So, it is hard to know what exactly any particular variable may capture. The authors are very clear that this is beyond the scope of their analysis. Nevertheless, I believe that results presented in the paper can collectively offer some suggestive evidence as to the possible channels behind this incredibly large explanatory power of 'no qualifications'.

Education (or the lack of it) could matter more for two broad reasons. First, people with low or no qualifications may be harder hit by immigration, globalization, and austerity measures. Part of education's strong explanatory power may thus be due to other economic factors left in the error term when considering demographic and education characteristics alone. Second, people with low or no qualifications may be exactly the demographic that is more easily swayed by the misinformation and unrealistic promises of the Leave campaign. Education may thus be proxying for the role of the anti-EU propaganda that dominated the tabloid press at the run-up to the election.

Results in the paper seem to suggest that although both explanations may be at work, the second explanation is probably more dominant. When all variables are allowed to compete against each other in the full model, the model's R^2 increases marginally to 0.88 as opposed to 0.77 when only demographic and education variables are included. Many of the coefficients of the demographic and education variables remain largely unchanged or even strengthen relative to their partial model. If their superior power was primarily due the electorate in those areas being harder hit by adverse shocks associated with immigration, globalization, and austerity measures, one would expect the education coefficients to lose economic and statistical significance in the full model when variables such as employment shares, wages, and fiscal cuts are included.

In addition, interaction terms between education and immigration or fiscal cut indicators yield mixed results. Consistent with the first explanation, positive and statistically significant coefficients are found between immigration from Eastern Europe and the share of unqualified people or share of people in manufacturing. However, the opposite is found for areas with a larger share of workers in Finance or higher median hourly pay. Instead, no clear pattern is found with respect immigration from EU 15 and non-EU countries or fiscal cuts. Interaction terms between share of the population with no qualifications and tabloid press penetration offer more decisive results indicating that the positive association between Vote Leave shares and 'no qualifications' is significantly stronger in areas with stronger tabloid penetration. Bearing in mind that readership of tabloid press may correlate with unobserved economic and demographic characteristics, these findings are consistent with the hypothesis that anti-EU propaganda from tabloid

press such as Daily Mail, Sun, and Express may indeed had a larger impact among the least educated.

Reassuringly, an out-of-sample analysis of the 2017 French Presidential election offers external validity to analysis indicating that the relations uncovered in the paper are not unique to the United Kingdom. The share of the population with no qualifications is also found to be associated with consistently larger vote shares for the far-right populist leader Marie Le Pen. (Perhaps the only notable difference between the two elections is that the unemployment rate seems to have played a more important role in the French election.) More broadly, the UK model explains a large fraction of the variation in the French vote (65%). An out-of-sample prediction of the French vote using the UK estimates also performs reasonably well. All in all, these results indicate that similar socio-economic factors were behind both elections and relations uncovered in the paper are not specific to the United Kingdom.

Andrea Mattozzi

European University Institute

Summary

This paper provides a detailed descriptive analysis of socio-economic characteristics that correlate with the vote Leave share in the UK June 2016 referendum. In particular, it shows that somewhat contrary to the popular wisdom, exposure to EU immigration and trade policy did not play a significant role in predicting a vote in favour of leaving the EU. On the other hand, education profiles, dependence on manufacturing employment, low income, and high unemployment appear to be correlated with support for Brexit. The paper also uncovers that the urban–rural divide, with voters living in urban areas less likely to vote for Brexit as opposed to voters in rural areas, is mostly the outcome of a composition effect. Turning attention to turnout data, the paper argues that even in the case in which younger voters – who were mostly in favour of Remain – would have turned out more, the outcome of the referendum would have been (likely) the same.

Comments

This is a very interesting paper that makes a valuable contribution in our understanding of the outcome of the recent UK referendum. In my opinion, the main weaknesses of the paper are four. First, very little can be said about the individual voting behaviour. This is clearly a data limitation and unfortunately not much can be done to overcome it. On the other hand, the lack of individual data somewhat limits the conclusion that we can draw from the analysis. Second, since the paper lacks a theoretical underpinning (which is not necessarily a problem per se), it is very hard to interpret and take something out from the counterfactuals on turnout, which remain mostly speculative. In

general I believe it can be dangerous to 'let the data speak for themselves'. Third, the paper dismisses, with not particularly strong arguments, a number of short-run narratives that might have had a non-negligible role in affecting the final outcome of the referendum, for example, the role of campaign and media exposure to affect both turnout and vote choice. In the last revision, the authors have made some steps to address these points. Finally, I am not sure what to take home in terms of policy implications. Regarding the last two points (turnout and policy implications), I have two comments.

First, while some of the very latest polls were predicting an advantage of the vote Leave supporters, it is hard to deny that the outcome came out as truly unexpected. This could be due to a number of reasons, but an important and often overlooked aspect is that polls do a pretty good job at predicting how people are going to vote, but do a poor job at predicting who will vote. And the reason is very simple: turnout is endogenous. Whether voters expect their party to win or lose changes whether they will bother to vote – so that voters' turnout is subject to some 'Uncertainty principle'. Let me give an example: Suppose that a pollster announce that Remain is going to win. Then a Leave supporter may choose to skip the vote all together. After all, if voting is costly and one side is expected to win then why bother to turn out? But then a Remain supporter, who realizes that turnout of the opposing side will be relatively low, might choose to skip the vote as well. As a result, a Leave supporter does not want to abstain any longer … etc. In a nutshell, the voters' turnout game has no pure strategy equilibrium, unless there is enough exogenous uncertainty. But then the reason why often pollsters have a hard time in predicting the election outcome is simply because turnout is hardly predictable. But there is more about endogenous turnout. If one considers the role that social norms, peer pressure, and monitoring play in voters' participation, it is not obvious that the larger party is always advantaged. Indeed, the age composition of Brexit supporters, their location, and turnout decisions could be consistent with a case of small party advantage in a peer pressure mobilization model. While the current paper does not need a formal model, a deeper link with voting theory might be good and group turnout models could be a good place to look at [see Levine and Mattozzi (2017) and references therein].

A second comment pertains the role played by referendums in aggregating disperse information. The bright side of direct democracy is the well-known Condorcet Jury Theorem: Even if information is dispersed, in the limit, the probability that a majority chooses the 'correct' alternative approaches 1 as the number of voters increases. Interestingly, however, if voting is about issues with distributional consequences and the identity of winners and losers is ex-ante uncertain, the Condorcet Jury Theorem may fail [see Bhattacharya (2013) and Ali *et al.* (2017)]. A natural question then arises: is the set of issues that are decided via a popular referendum exogenous? Not really. Often referendums are called on divisive policy issues over which elected politicians are not able and willing to take a clear position and act accordingly. In this respect, issues with distributional consequences and such that the identity of winners and losers is ex ante uncertain to ascertain seem very good candidates for information aggregation to

potentially fail. After all we should not be too surprised by 'weird' referendum outcomes, and be suspicious when direct democracy is advocated by populist movements.

SUPPLEMENTARY DATA

Supplementary data are available at *Economic Policy* online.

REFERENCES

Ali, S.N., M. Mihm and L. Siga (2017). 'The perverse politics of polarization', Working Paper. Available at: https://nageebali.wordpress.com/research/.

Arnorsson, A. and G. Zoega (2016). 'On the causes of Brexit', CESifo Working Paper 6056.

Arzheimer, K. (2009). 'Contextual factors and the extreme right vote in Western Europe, 1980–2002', *American Journal of Political Science*, 53, 259–75.

Ashcroft, M. (2016). How the United Kingdom Voted on Thursday. . . and Why. lordashcroft.com, June 2016 accessed 30 June 2016.

Autor, D., D. Dorn, G. Hanson and K. Majlesi (2016). 'Importing political polarization? The electoral consequences of rising trade exposure', NBER Working Paper 22637.

Barone, G., A. D'Ignazio, G. de Blasio and P. Naticchioni (2016). 'Mr. Rossi, Mr. Hu and Politics. The role of immigration in shaping natives' voting behavior', *Journal of Public Economics*, 136, 1–13.

Beatty, C. and S. Fothergill (2016). 'Hitting the poorest places hardest', Technical report, Centre for Regional Economic and Social Research, Sheffield Hallam University.

Becker, S., P. Egger and M. von Ehrlich (2010). 'Going NUTS: the effect of EU structural funds on regional performance', *Journal of Public Economics*, 94, 578–90.

Becker, S.O., P.H. Egger and M. von Ehrlich (2012). 'Too much of a good thing? On the growth effects of the EU's regional policy', *European Economic Review*, 56, 648–68.

——. (2013). 'Absorptive capacity and the growth and investment effects of regional transfers: a regression discontinuity design with heterogeneous treatment effects', *American Economic Journal: Economic Policy*, 5, 29–77.

Becker, S.O. and T. Fetzer (2016). 'Does migration cause extreme voting?', CAGE Working Paper 306 (October).

Becker, S.O., T. Fetzer and D. Novy (2016). 'Who voted for Brexit? A comprehensive district-level analysis', CAGE Online Working Paper Series 305, Competitive Advantage in the Global Economy (CAGE).

Bell, B. and S. Machin (2016). Brexit and wage inequality. VoxEU, August 2016.

Bhattacharya, S. (2013). 'Condorcet Jury theorem in a spatial model of elections', Working Paper. Available at: https://www.sites.google.com/site/souravgati/.

Borjas, G. (2003). 'The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market', *Quarterly Journal of Economics*, 118, 1335–74.

Borjas, G. and J. Monras (2016). 'The labor market consequences of refugee supply shocks', NBER Working Paper 22656.

Burgoon, B. (2012). 'Inequality and anti-globalization backlash by political parties', *European Union Politics*, 14, 408–35.

Burn-Murdoch, J. (2016a). Brexit: everything you wanted to know about turnout by age at the EU referendum. *Financial Times*, 01 July 2016.

——. (2016b). Brexit: voter turnout by age. *Financial Times*, 24 June 2016.

Butler, D. and U. Kitzinger (1976). *The 1975 Referendum*, Palgrave Macmillan, London.

Casella, A. and A. Gelman (2008). 'A simple scheme to improve the efficiency of referenda', *Journal of Public Economics*, 92, 2240–61.

Clarke, H., M. Goodwin and P. Whiteley (2016). 'Leave was always in the lead: why the polls got the referendum result wrong', *British Politics and Policy Blog, London School of Economics and Political Science*, July 2016.

Clarke, H., P. Whiteley, W. Borges, D. Sanders and M. Stewart (2016). 'Modelling the dynamics of support for a right-wing populist party: the case of UKIP', *Journal of Elections, Public Opinion and Parties*, 26, 134–54.

Colantone, I. and P. Stanig (2016). 'Global competition and Brexit', Baffi Carefin Centre Research Paper Series No. 2016-44, September 2016.
Cortes, P. (2008). 'The effect of low-skilled immigration on U.S. prices: evidence from CPI data', *Journal of Political Economy*, 116, 381–422.
Crafts, N. (2016). 'The growth effects of EU membership for the UK: a review of the evidence', CAGE Working Paper 280.
Curtice, J. (2016). Brexit buyers' remorse? There's more evidence of abstainers' regret. *British Politics and Policy Blog, London School of Economics and Political Science*, July 2016.
Davies, W. (2016). 'Thoughts on the Sociology of Brexit', Blog, Political Economy Research Centre, Goldsmiths, University of London.
DellaVigna, S., J. List, U. Malmendier and G. Rao (2017). 'Voting to tell others', *Review of Economic Studies*, 84, 143–81.
Dippel, C., R. Gold and S. Heblich (2015). 'Globalization and its (dis-)content: trade shocks and voting behavior', NBER Working Paper 21812.
Dorling, D. (2016). 'Brexit: the decision of a divided country', *BMJ*, 354.
Druckman, J. and A. Lupia (2000). 'Preference formation', *Annual Review of Political Science*, 3, 1–24.
Eichengreen, B. (2008). *The European Economy Since 1945: Coordinated Capitalism and Beyond*, Princeton University Press, Princeton, NJ.
Evans, G. (2000). 'The continued significance of class voting', *Annual Review of Political Science*, 3, 401–17.
Ferree, K., G. B. Powell and E. Scheiner (2014). 'Context, electoral rules, and party systems', *Annual Review of Political Science*, 17, 421–39.
Findlay, R. and K. O'Rourke (2007). *Power and Plenty: Trade, War, and the World Economy in the Second Millennium*, Princeton University Press, Princeton, NJ.
Ford, R. and M. Goodwin (2014). *Revolt on the Right: Explaining Support for the Radical Right in Britain*, Routledge, Abingdon, Oxfordshire.
Funk, P. and C. Gathmann (2015). 'Gender gaps in policy making: evidence from direct democracy in Switzerland', *Economic Policy*, 30, 141–81.
Gerber, A., D. Green and C. Larimer (2008). 'Social pressure and voter turnout: evidence from a large-scale field experiment', *American Political Science Review*, 102, 33–48.
Goodwin, M. and O. Heath (2016). 'The 2016 referendum, Brexit and the left behind: an aggregate-level analysis of the result', *The Political Quarterly*, 87, 323–32.
Guiso, L., P. Sapienza and L. Zingales (2016). 'Monnet's error?', *Economic Policy*, 31, 247–97.
Guyon, I. and A. Elisseeff (2003). 'An introduction to variable and feature selection', *Journal of Machine Learning Research*, 3, 1157–82.
Hastie, T., R. Tibshirani and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, Springer, New York, NY.
Helm, T. (2016). EU referendum: youth turnout almost twice as high as first thought. TheGuardian.com, 10 July 2016 accessed 10 July 2016.
Herrera, H. and A. Mattozzi (2010). 'Quorum and turnout in referenda', *Journal of the European Economic Association*, 8, 838–71.
Innes, D. and G. Tetlow (2015). 'Delivering fiscal squeeze by cutting local government spending', *Fiscal Studies*, 36, 303–25.
Jennings, W. and G. Stoker (2016). 'The bifurcation of politics: two Englands', *The Political Quarterly*, 87, 372–82.
Kaufmann, E. (2016). 'It's NOT the economy, stupid: Brexit as a story of personal values', British Politics and Policy Blog, London School of Economics and Political Science, July 2016.
Kinsman, J. (2016). A Brexit post-mortem: 17 takeaways for a fallen David Cameron. *OpenCanada.org*, 04 July 2016.
Levine, D. and A. Mattozzi (2017). 'Voter turnout with peer punishment', Working Paper, March 2017. Available at: http://apps.eui.eu/Personal/Amattozzi/Papers/Group%20Voting.pdf.
Lewis-Beck, M. and M. Stegmaier (2000). 'Economic determinants of electoral outcomes', *Annual Review of Political Science*, 3, 183–219.
Los, B., P. McCann, J. Springford and M. Thissen (2017). 'A mismatch between local voting and the local economic consequences of Brexit', *Regional Studies*, 51, 786–99.

Madestam, A., D. Shoag, S. Veuger and D. Yanagizawa-Drott (2013). 'Do political protests work? Evidence from the tea party movement', *Quarterly Journal of Economics*, 128, 1633–85.
Matsusaka, J.G. (1992). 'Economics of direct legislation', *Quarterly Journal of Economics*, 107, 541–71.
Meier, A., L. Schmid and A. Stutzer (2016). 'Rain, emotions and voting for the status quo', IZA Discussion Paper 10350.
Mullainathan, S. and J. Spiess (2017). 'Machine learning: an applied econometric approach', *Journal of Economic Literature*, 31, 87–106.
Müller, J. (2016). *What is Populism?* University of Pennsylvania Press, Philadelphia, PA.
O'Rourke, K. (2016). Brexit: this backlash has been a long time coming. *VoxEU*, August 2016.
Rodrik, D. (2016). Brexit and the globalization trilemma. *Dani Rodrik's Weblog*, 13 June 2016.
Rothwell, J. and P. Diego-Rosell (2016). 'Explaining nationalist political views: the case of Donald Trump', Gallup Working Paper, November 2016.
Vargas-Silva, C. (2012). EU migrants in other EU countries: an analysis of bilateral migrant stocks. *Briefing*, the Migration Observatory at the University of Oxford.
Wadsworth, J., S. Dhingra, G. Ottaviano and J. Van Reenen (2016). 'Brexit and the impact of immigration on the UK', Brexit Analysis No. 5, Centre for Economic Performance, London School of Economics and Political Science.
Whitaker, R. and P. Lynch (2011). 'Explaining support for the UK independence party at the 2009 European parliament elections', *Journal of Elections, Public Opinion and Parties*, 21, 359–79.