WARWICK
THE UNIVERSITY OF WARWICK

Manuscript version: Working paper (or pre-print)
The version presented here is a Working Paper (or 'pre-print') that may be later published elsewhere.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/107728

How to cite:
Please refer to the repository item page, detailed above, for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

warwick.ac.uk/lib-publications

WARWICK
ECONOMICS

Who Voted for Brexit?
Individual and Regional Data Combined

Eleonora Alabrese, Sascha O. Becker,
Thiemo Fetzer and Dennis Novy

(This paper also appears as CAGE discussion paper 384)

August 2018 No: 1172

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

Who Voted for Brexit?
Individual and Regional Data Combined

Eleonora Alabrese Sascha O. Becker
Thiemo Fetzer Dennis Novy*

August 2018

Abstract

Previous analyses of the 2016 Brexit referendum used region-level data or small samples based on polling data. The former might be subject to ecological fallacy and the latter might suffer from small-sample bias. We use individual-level data on thousands of respondents in *Understanding Society*, the UK's largest household survey, which includes the EU referendum question. We find that voting Leave is associated with older age, white ethnicity, low educational attainment, infrequent use of smartphones and the internet, receiving benefits, adverse health and low life satisfaction. These results coincide with corresponding patterns at the aggregate level of voting areas. We therefore do not find evidence of ecological fallacy. In addition, we show that prediction accuracy is geographically heterogeneous across UK regions, with strongly pro-Leave and strongly pro-Remain areas easier to predict. We also show that among individuals with similar socioeconomic characteristics, Labour supporters are more likely to support Remain while Conservative supporters are more likely to support Leave.

Keywords: Aggregation, Ecological Fallacy, European Union, Populism, Referendum, UK
JEL Classification: D72, I10, N44, R20, Z13

Acknowledgements: We are grateful to the editor and two referees for constructive comments. We thank the *Understanding Society* team based at the Institute for Social and Economic Research at the University of Essex for access and guidance on the data. All authors are based at the University of Warwick. Becker, Fetzer and Novy are affiliated with the Centre for Competitive Advantage in the Global Economy (CAGE). Becker is also affiliated with CEPR, CESifo, ifo, IZA and ROA. Fetzer is also affiliated with SERC. Novy is also affiliated with CEPR, CESifo and CEP/LSE. *Email addresses:* e.alabrese@warwick.ac.uk, s.o.becker@warwick.ac.uk, t.fetzer@warwick.ac.uk, d.novy@warwick.ac.uk.

1 Introduction

Populism has been on the rise across Europe and the United States in recent years, culminating in the election of Donald Trump as US President and the Brexit vote in the 2016 EU referendum. The Brexit vote came as a shock to many observers and triggered early attempts to understand the voting patterns.[1] These studies relied almost exclusively on aggregate data at the level of voting areas. Regressing vote shares across voting areas on average population characteristics risks falling into the ecological fallacy trap of inferring *individual* associations from *aggregate* data (see Robinson, 1950).

We use detailed individual-level data from the *Understanding Society* survey containing the EU referendum question to address three interrelated questions. First, we investigate the relationship between voters' personal characteristics and their expressed voting intentions. Particularly, we address whether ecological fallacy may be driving the associations documented in the aggregated data. Second, building a predictive model of Leave support we assess which voting determinants have the most power to predict voting behavior out of sample. Third, we investigate the classification errors that this predictive model makes by region and voters' closeness to political parties.

We find that individual and aggregate coefficients point in a similar direction, suggesting that ecological fallacy is of limited concern. Second, we document that the predictive models exhibit a significant gain in accuracy when exploiting both individual and regional variables. Lastly, we document that a predictive model performs best in parts of the UK with the most extreme referendum outcomes: Lincolnshire (highest Leave share) and London (lowest Leave share across mainland Britain). Furthermore, a decomposition of classification errors reveals that closeness to a political party is likely an important omitted variable, suggesting that unobservable traits and identity are further key correlates.

The paper is structured as follows. Section 2 lays out the literature background, describes the data and explains our empirical approach. We present

[1] See Burn-Murdoch (2016) in the Financial Times as an example of various correlation plots; more in-depth work followed, for example Clarke and Whittaker (2016), Darvas (2016), Langella and Manning (2016).

graphical summaries of our results in Section 3, and we conclude in Section 4. Underlying regression results and further details are relegated to an appendix.

2 Background, data and empirical approach

2.1 Background

This paper builds on Becker et al. (2017) who analyze the Brexit vote shares across UK voting areas, using a wide range of explanatory variables. They show that the Leave vote shares are systematically correlated with older age, lower educational attainment, unemployment, or employment in certain industries such as manufacturing, as well as with a lack of quality of public service provision.

These results fit in with other evidence on the Brexit vote. An early attempt to explain the referendum outcome was made by Ashcroft (2016) whose polling data indicated that the typical Leave voter is white, middle class and lives in the South of England. Sampson (2017) reviews the literature on the likely economic consequences of Brexit on the British economy and other countries.

Our paper also relates to the wider literature on political polarization as well on voting for far-right parties. Ferree et al. (2014) provide an extensive review of academic works which link voting patterns to demographic, economic and political features. Voters' behaviour has also been shown to be strongly associated with individual scepticism towards institutions (e.g. Euroscepticism) or intolerance against foreigners (see Whitaker and Lynch, 2011; Clarke et al., 2016 and Arzheimer, 2009). Additional studies claim that ethnic minorities may engage in 'ethnic' or 'policy' voting depending on the issue they are called to vote upon (see Bratton and Kimenyi, 2008 and Tolbert and Hero, 1996).

Polarization has also been related to immigration (see Barone et al., 2016) as well as trade integration (Dippel et al., 2015; Burgoon, 2012 and Autor et al., 2016). In the UK context, Becker and Fetzer (2016) examine immigration from Eastern Europe as a potential driver of support for the UK Independence Party, while Fetzer (2018) explores the role of austerity policies since 2010.

Overall, the voting patterns in the Brexit referendum are complex. One pos-

sible – albeit not the only – interpretation of the empirical literature on Brexit so far is that some people who favor Leave may feel 'left behind', be it economically or culturally (see Hobolt, 2016 and Clarke et al., 2017). This is consistent with sociological studies which demonstrate similar patterns for the Tea Party Movement and the 2016 US presidential election, e.g. Hochschild (2016).

2.2 Data

Are these aggregate patterns found by Becker et al. (2017) and others a fair reflection of individual-level relationships? The individual-level data from wave 8 of the *Understanding Society* survey makes it possible to investigate this question. Our focus is on individual socio-economic variables for which region-level equivalents are used in Becker et al. (2017). Our approach of combining individual-level and aggregate data allows us (a) to check whether ecological fallacy is an important factor in aggregate analyses of the Brexit vote, and (b) to exploit the combined predictive power of individual-level and aggregate variables. This opens up insights into (c) geographic heterogeneity in predictive power across UK regions.

The *Understanding Society* data cover a wide range of topics, in particular basic demographic data for all household members such as sex, age and ethnicity, place of birth, family background including marital status, educational attainment, current job characteristics, housing characteristics (owning vs. renting), health status and life satisfaction. We describe the sampling design in more detail in an appendix, and how we construct our sample (also see Knies, 2016).

2.3 Descriptive statistics

According to the summary statistics in Table 1, 42.2% of the 13,136 individuals in our sample indicate that the UK should leave the EU in response to the survey question "Should the UK remain a member of the EU or leave the EU?" This compares to 51.9% of the electorate voting Leave in the referendum. We refer to Becker et al. (2017, section 3.1) for a discussion of the aggregate voting and turnout patterns in the 2016 referendum.

As for demographics, the proportion of males is 45.4% of all individu-

als in the sample, while just about three out of ten respondents are aged 60 or above. People with no qualification account for about 8% of the sample. Roughly 90% of respondents are born in the UK. Asians are the largest ethnic minority amounting to 5.8% of the sample, followed by blacks (2.5%).[2] Over half of respondents are married or in a civil partnership. In terms of current employment, roughly four out of ten people declare to be without a paid job or to not have worked in the seven days prior to being questioned.[3]

2.4 Understanding Society: Research in progress

We gained access to *Understanding Society* data in the summer of 2017, at the same time as other groups of researchers in a pilot 'early access' project. We briefly summarize related preliminary findings reported by other researchers in short presentations in the summer of 2017. For instance, Creighton and Amaney (2017) find that opposition to immigration played a key role. Martin and Sobolewska (2017) explore racial determinants and find that ethnic minorities are strongly in favor of remaining in the EU. De Vries and Solaz (2017) attempt to explain voters' behavior by analyzing socio-economic determinants such as asset holdings, sources of income and skills, whereas Doebler et al. (2017) explore additional potential drivers such as personal economic struggle and regional economic decline.

As far as we are aware, only one other paper using *Understanding Society* data has come out as a working paper so far. Liberini et al. (2017) show that individuals dissatisfied with their own financial situation were more likely to vote Leave and that the very young were most likely to vote Remain. In related work, Pollock (2017) uses the Innovation Panel to argue that the rise in populism and the vote in favor of Brexit can be attributed to generational shifts away from mainstream political parties over the past three decades.

[2]Note that we sourced nationality and ethnicity variables also from earlier waves.

[3]The aggregate variables in Table 1 are not standardized for descriptive purposes, but they are in all regressions.

2.5 Empirical approach

We start with a simple model where the dependent variable y_{ic} is a dummy for individual i in local authority c which takes on the value 1 if the interviewed person answers "Leave" in response to the question "Should the UK remain a member of the EU or leave the EU?" and 0 if the answer is "Remain":

$$y_{ic} = \mathbf{x}'_{ic}\beta + \mathbf{z}'_{c}\delta + \varepsilon_{ic}. \tag{1}$$

The independent variables in the model are the *Understanding Society* cross-sectional individual covariates $\mathbf{x}_{ic}$ on the one hand, and area-specific aggregate variables $\mathbf{z}_c$ from Becker et al. (2017) on the other. Our overall sample contains 13,136 respondents for our baseline regressions. We also analyze smaller samples and subgroups of variables since not all *Understanding Society* respondents were asked each survey module. As the summary statistics in Table 1 show, roughly 42% of respondents are in favor of Leave.

We relegate the details of the underlying regression results to the appendix. For ease of interpretation, throughout the regression tables in the appendix we provide coefficients obtained from a simple linear probability model estimation of equation (1). However, each model is also estimated using the corresponding logistic regression model to provide an estimate of the success rate at the bottom of each table.

Since our interest centers on prediction, we need a metric to assess predictive accuracy of our regression models. We perform a simple validation exercise known from the machine learning literature. Our sample is divided into a random *training set* (2/3 of the sample) and a *validation set*. Logistic regressions are conducted on the training set, and we use the validation set to perform classification. We follow Bayes' optimal decision rule and classify an observation as "Leave" if the predicted posterior probability exceeds 50%. In essence, this rule simple allocates the label ("Leave" or "Remain") to an observation that, conditional on our predictors/features, is most likely. This decision rule minimizes the error rate or maximizes overall accuracy. Yet, it does so putting an equal penality or cost on false-positives versus false-negatives. The comparison of the predicted to the actual assignments allows us to estimate the out-of-sample

predictive power and to shed light on the two types of prediction errors (false positives versus false negatives). For instance, individual A in the validation set may, based on her characteristics, look like a typical Remain voter but is in reality a Leave voter, so we have a case of a false negative. Individual B in the validation set may, based on her characteristics, look like a typical Leave voter but is in reality a Remain voter, so we have a case of a false positive.

We stress that causality is beyond the scope of our paper. Instead, our results reflect a broad range of correlation patterns relating voting intentions to fundamental socio-economic features.[4] In our earlier work (Becker et al., 2017), we grouped variables by four topics: (1) EU exposure: immigration, trade and EU transfers; (2) Public service provision and fiscal consolidation; (3) Demography, education and life satisfaction; (4) Economic structure, wages and unemployment. Those groupings follow from prominent hypotheses that have been proposed to explain the EU referendum result. That is, the first grouping looks at the relationship between EU exposure and Leave voting. Here, we follow the same logic and look at groups of variables that correspond to one specific set of explanations for the referendum result. For each variable grouping, we assess its predictive power by itself, and compare this to the joint predictive power of all groups of variables combined. We discuss the different groupings in more detail in the appendix (the regression tables using the groups of variables under discussion are described in sections A.4-A.10 in the appendix).[5]

As Becker et al. (2017) explain, the fundamental difference between prediction, as pursued in this paper, and causal inference is as follows. Causal inference focuses on the internal validity of causally estimated reduced-form (or structural) parameters β. In contrast, prediction is concerned with the external validity of the estimated fitted values $\hat{y}$.[6] Causal inference seeks to obtain a set

[4] In a fascinating paper, Colantone and Stanig (2018) focus on one specific causal factor behind the Leave vote: rising import competition from China. While papers studying causality are extremely important, they give prominence to one factor at a time, an aim different from ours which is to look at the relative predictive power of different variables.

[5] One might wonder whether including region fixed effects above and beyond the individual-level and region-level predictors is beneficial in terms of predition accuracy, but the benefits are very marginal in our case. Since region fixed effects are a 'black box', we refrain from including them given the very limited gains.

[6] While we do not use machine-learning methods in this paper such as best subset selection (BSS) or LASSO, we did so in Becker et al. (2017), i.e. our selection of variables is guided by the (aggregate) variables employed in that earlier paper.

of estimated parameters $\hat{\beta}$ that are usually studied in isolation. Thus, they often do not render themselves useful for prediction because 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). The underlying estimated parameters that yield good model fit are typically of limited interest per se. For this reason, we only show coefficient estimates in appendix tables, while in the main text we focus on graphical representation.

3 Predicting the vote

In order to focus on prediction quality, we relegate the discussion of individual regression tables to the appendix. First, we focus on the relative predictive power of individual-level and aggregate variables. Second, we examine the predictive power of our best-performing model across regions and lastly, we investigate the classification error structure.

3.1 Individual vs. aggregate variables

Figure 1 reports the proportion of correct predictions (success rates) for each variable grouping estimated in the hold-out sample. In particular, Figure 1(a) illustrates success rates for (groupings of) aggregate variables and Figure 1(b) for individual-level variables. Figure 1(c) combines aggregate and individual-level variables. Figure 1(d) reports success rates for non-comparable individual variables.

The overall classification success rate when relying on aggregate data in Figure 1(a) is 58.8%. In the narrow individual-level sample for which employment and related individual data is collected in the *Understanding Society* sample, the overall accuracy reaches 62.9% using the aggregate level area employment characteristics. The improvement in terms of accuracy relative to a naive classification rule that classifies everyone as Remain (generating a success rate of 57.8%, i.e. one minus the sample 'Leave' share) thus is only modest. When focusing on all comparable individual-level covariates in Figure 1(b), we see that

individual-level variables have stronger predictive power than aggregate ones. The improvement in accuracy up to 63.4% with all variables included suggests an improvement in prediction accuracy relative to the naive benchmark of 9.7%.

Furthermore, an inspection of the tables in the appendix confirms that the individual-level predictors yield broadly similar sign patterns to their aggregate level equivalents. This suggests that ecological fallacy is not a major concern for the results in Becker et al. (2017).

The combination of individual and aggregate characteristics yields a further slight improvement in prediction accuracy. Relative to the naive classification rule, accuracy can improve up to 64.6% with all covariates included, representing an improvement of 11.7% in relative terms. Adding further individual-level characteristics that are included in the *Understanding Society* sample (but for which no aggregate proxy measures exist) suggests that overall accuracy is not further improved.

In fact, our best model including all characteristics sees a small drop in the success rate. In terms of the bias-variance trade-off inherent in such predictive models, the improvement in terms of bias are therefore likely offset by an inflation in terms of variance, resulting in worse out-of-sample performance. We refer to James, Witten, Hastie and Tibshirani (2013) for a discussion of the bias-variance trade-off.

As explained in the appendix, we explore a number of novel individual determinants. We find that marital status, technology use and dependence on income support and state benefits are all systematically linked to individual voting behavior. In particular, individuals who do not possess smartphones and who use the internet infrequently appear more inclined to support Leave. Those repeatedly seeking health care or receiving income support also tend to be more in favor of Brexit. Similarly, it is also fair to say that Brexit is a predominantly white phenomenon compared to ethnic minorities.

3.2 Geographical heterogeneity

An instructive step lies in attempting to decompose in which regions our model does a good job in correctly classifying the voting intentions in the *Understanding*

Society sample. Among all NUTS2 regions in Figure 2, Inner London displays the lowest error rate (21%) followed by Lincolnshire and North Eastern Scotland (with 23% and 26%, respectively). Lincolnshire and Inner London had among the highest and lowest Leave vote shares in the referendum. Thus, it is hardly surprising that the empirical model performs well in separating voters in these regions.

The model has the lowest performance in Tees Valley and Durham, East Anglia, and Merseyside (with error rates around 43-44%). Generally, the picture that emerges suggests that purely based on the socio-economic characteristics, areas that are more disadvantaged are the ones where it is most difficult to separate Leave from Remain voters. Non-economic factors may therefore be particularly helpful in capturing variation between voters in these areas.

3.3 Types of errors

We turn to decomposing errors into false positives and false negatives. The results presented in Figure 2 suggest that the regions of Inner and Outer London, Berkshire, Buckinghamshire and Oxford as well as North Eastern Scotland stand out as having the highest rate of false negatives (blue bars). False negatives are cases in which our model identifies an individual as a Remain voter, while in fact they state an intention to vote Leave. The false negatives in Figure 2 suggest that there are non-negligible proportions of voters who, based on their socio-economic characteristics, *look like Remain voters* but actually express an intention to vote Leave. In Outer London, 80% of all classification errors are false negatives. The same holds true for many of the other regions in London's wealthy commuter belt.

We next investigate whether classification errors can be related to individual political party preferences. From previous *Understanding Society* survey rounds which asked participants what party they felt closest to, we obtain that measure for 65% of our estimation sample. Figure 3 highlights that, while overall accuracy across the stated historical party preferences is similar, the type of classification error is quite heterogeneously distributed. In particular, Labour voters are more likely to contribute to the false positive errors – cases where our

model classifies an individual as a Leave voter when in fact they favour Remain – making up 51.27% of the share of all false positives. By contrast, Conservative party supporters make up 44.8% of the share of false negatives – individuals who look like Remain voters but actually intend to vote Leave.

Overall, our findings indicate that Labour voters with observables that put them in the Leave camp – male, older, less educated, less likely to be in employment, etc. – are significantly more likely to express a preference for the status quo of remaining in the EU. Voters with similar socio-economic profiles who identify with the Conservative party are more likely to vote Leave. This suggests the potential importance of other characteristics not in the data set, for instance psychological traits such as openness as well as attitudes towards national identity.

4 Conclusion

Individual-level regressors from the British *Understanding Society* survey containing the 2016 EU referendum question give similar results to corresponding aggregate variables at the level of local authority areas analyzed by Becker et al. (2017). We therefore find no evidence of ecological fallacy effects – individuals appear to behave in similar ways as suggested by the aggregate data.

We also shed light on the predictive power of different determinants of the Leave vote. Demographics and employment characteristics are the most relevant covariates for prediction, while the cumulative power of individual-level and aggregate variables shows a non-negligible gain over aggregate data alone. Geographical heterogeneity is also important as our model performs best in more prosperous areas (London in particular).

Finally, we also find that individuals who support the Labour party but have otherwise observables that would put them in the Leave camp are significantly more likely to vote Remain. Vice versa, supporters of the Conservative party with Remain-favouring characteristics are more likely to vote Leave.

References

Arzheimer, K. (2009). Contextual Factors and the Extreme Right Vote in Western Europe, 1980-2002. American Journal of Political Science 53(2): 259–275.

Ashcroft, M. (2016). How the United Kingdom Voted on Thursday... and Why. Lordashcroft.com, 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.

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 (2017). Who Voted for Brexit? A Comprehensive District-Level Analysis. Economic Policy 32(92): 601–650.

Bratton, M. and M. S. Kimenyi (2008). Voting in Kenya: Putting Ethnicity in Perspective. Journal of Eastern African Studies 2(2): 272–289.

Buck, N. and S. McFall (2012). Understanding Society: Design Overview. Longitudinal and Life Course Studies 3(1): 5–17.

Burgoon, B. (2012). Inequality and Anti-Globalization Backlash by Political Parties. European Union Politics 14(3): 408–435.

Burn-Murdoch, J. (2016). Brexit: Voter Turnout by Age. Financial Times, 24 June 2016.

Clarke, H., M. Goodwin and P. Whiteley (2017). Why Britain Voted for Brexit: An Individual-Level Analysis of the 2016 Referendum Vote. Parliamentary Affairs 70(3): 439–464.

Clarke, S. and M. Whittaker (2016). The Importance of Place: Explaining the Characteristics Underpinning the Brexit Vote Across Different Parts of the UK. London: Resolution Foundation. *(http://www.resolutionfoundation.org/publications/the-important-of-place-explaining-thecharacteristics-underpinning-the-brexit-vote-across-different-parts-of-the-uk/)*

Colantone, I. and P. Stanig (2018). Global Competition and Brexit. American Political Science Review 112(2): 201–218.

Creighton, M. J. and J. Amaney (2017). Masked Intolerance toward Immigrants in the UK before and after the Vote to Leave the EU. Understanding Society Scientific Conference. July 2017, University of Essex. Unpublished conference paper.

Darvas, Z. (2016). High Inequality and Poverty Helped Trigger the Brexit Protest Vote. London: London School of Economics and Political Science.
(http://blogs.lse.ac.uk/brexit/2016/08/31/brexit-should-be-a-wake-up-call-in-the-fight-against-inequality/)

De Vries, C. and H. Solaz (2017). Brexit and the Left Behind? Disentangling the Effects of Assets, Benefits, Income and Skills on Voting Behaviour in the 2016 Brexit Referendum. EU Referendum Project Symposium. June 2017, University of Essex. Unpublished conference paper.

Dippel, C., R. Gold and S. Heblich (2015). Globalization and Its (Dis-)Content: Trade Shocks and Voting Behavior. NBER Working Paper 21812.

Doebler, S., G. Catney, S. Shortall and B. Hayes (2017). What is the Role of Regional Economic Decline, Personal Economic Struggle and Political Disengagement as Drivers of Support for Brexit? A Longitudinal Analysis using Understanding Society (UKHLS) Data. EU Referendum Project Symposium. June 2017, University of Essex. Unpublished conference paper.

Ferree, K., G. B. Powell and E. Scheiner (2014). Context, Electoral Rules, and Party Systems. Annual Review of Political Science 17: 421–439.

Fetzer, T. (2018). Did Austerity Cause Brexit? CAGE Working Paper.

Hobolt, S. (2016). The Brexit Vote: A Divided Nation, A Divided Continent. Journal of European Public Policy 23(9): 1259–1277.

Hochschild, A. R. (2016). Strangers in Their Own Land: Anger and Mourning on the American Right. The New Press, New York.

James, G., D. Witten, T. Hastie and R. Tibshirani (2013). An Introduction to Statistical Learning: With Applications in R. Springer, New York.

Knies, G. (2016). Understanding Society - the UK Household Longitudinal Study: Waves 1-6, (2009-2015) User Guide. University of Essex, Colchester.

Langella, M. and A. Manning (2016). Who Voted Leave: The Characteristics of Individuals Mattered, But So Did Those of Local Areas. London: London School of Economics and Political Science. *(http://blogs.lse.ac.uk/politicsandpolicy/explaining-the-vote-for-brexit/).*

Liberini, F., A. J. Oswald, E. Proto and M. Redoano (2017). Was Brexit Caused by the Unhappy and the Old? IZA Discussion Paper No. 11059

Martin, N. and M. Sobolewska (2017). Ethnic Minority Support for Leaving the European Union. EU Referendum Project Symposium. June 2017, University of Essex. Unpublished conference paper.

Mullainathan, S. and J. Spiess (2017). Machine Learning: An Applied Econometric Approach. Journal of Economic Literature 31(2): 87–106.

Pollock, G. (2017). Melancholy Politics: Brexit as a Consequence of Weakening Mainstream Party Attachment in the United Kingdom. Understanding Society Scientific Conference. July 2017, University of Essex. Unpublished conference paper.

Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. American Sociological Review 15(3): 351–357.

Sampson, T. (2017). Brexit: The Economics of International Disintegration. Journal of Economic Perspectives 31(4): 163–184.

Tolbert, C. J. and R. E. Hero (1996). Race/Ethnicity and Direct Democracy: An Analysis of California's Illegal Immigration Initiative. The Journal of Politics 58(3): 806–818.

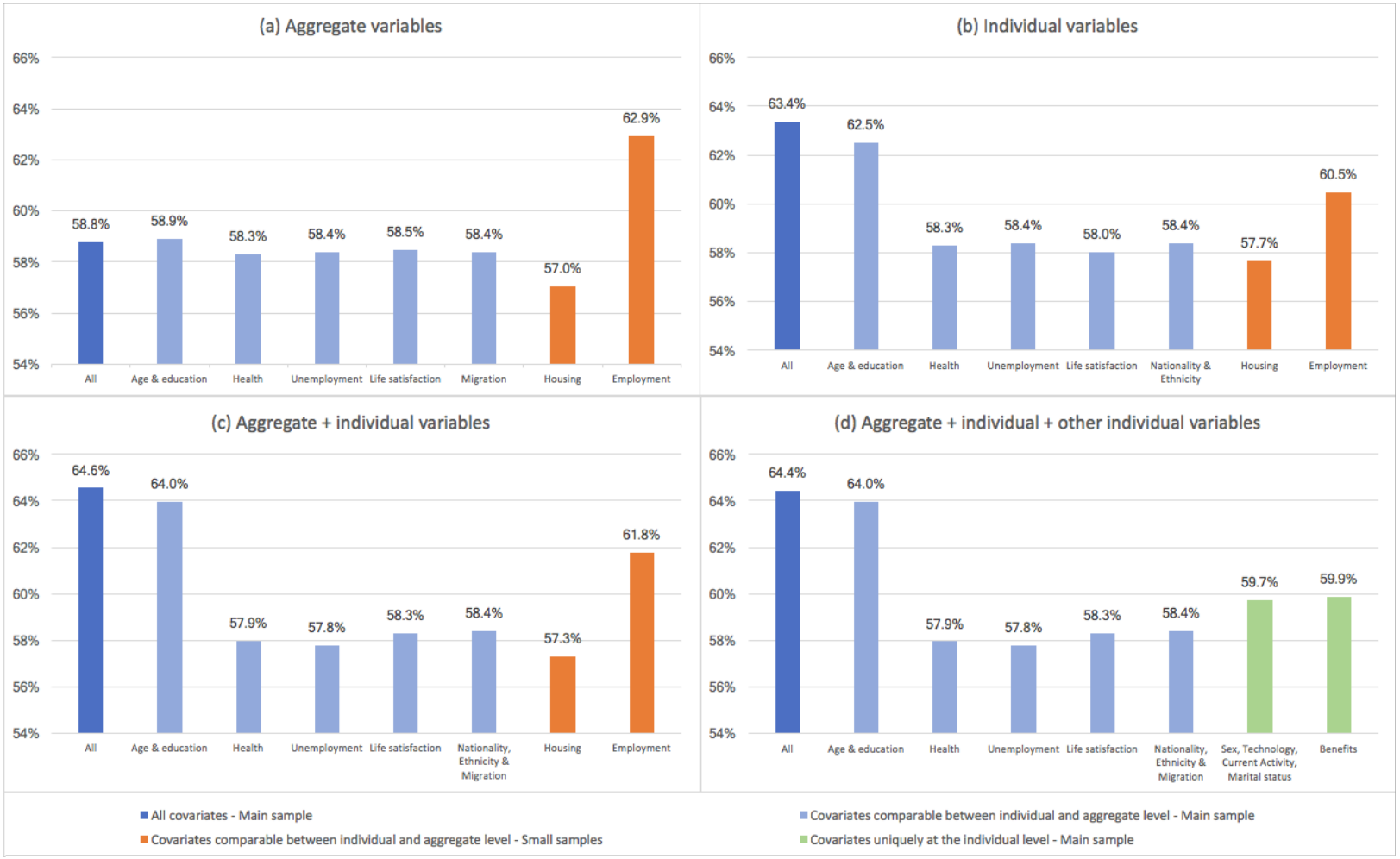
Whitaker, R. and P. Lynch (2011). Explaining Support for the UK Independence Party at the 2009 European Parliament Elections. Journal of Elections, Public Opinion and Parties 21(3): 359–379.

Table 1: Summary statistics

VARIABLES	N	mean	sd	min	max
Dependent variable:					
Should the UK leave the EU	13,136	0.422	0.494	0	1
Individual variables:					
Sex = Male	13,136	0.454	0.498	0	1
Age = 60 or older	13,136	0.305	0.460	0	1
Highest qualification = Other lower qualification	13,136	0.0847	0.278	0	1
Highest qualification = No qualification	13,136	0.0796	0.271	0	1
Frequency using internet = Every day	13,136	0.792	0.406	0	1
Frequency using internet = No access	13,136	0.0153	0.123	0	1
Born in UK	13,136	0.905	0.294	0	1
Ethnic group = White	13,136	0.896	0.306	0	1
Ethnic group = Asian	13,136	0.0579	0.234	0	1
Ethnic group = Black	13,136	0.0245	0.155	0	1
Current legal marital status = Single	13,136	0.287	0.452	0	1
Current legal marital status = Married or civil partner	13,136	0.546	0.498	0	1
Visits GP in 12m = None	13,136	0.214	0.410	0	1
Visits GP in 12m = Over 10	13,136	0.0613	0.240	0	1
Housing tenure = Owned (outright + mortgage)	9,344	0.664	0.472	0	1
No work last week & doesn't have paid job	13,136	0.379	0.485	0	1
Current job sector = Manufacturing	7,950	0.0826	0.275	0	1
Income support	13,136	0.0158	0.125	0	1
Dissatisfied with health	13,136	0.244	0.430	0	1
Dissatisfied with income	13,136	0.214	0.410	0	1
Aggregated variables:					
Unemployment rate (2015)	13,136	5.618	2.159	1.600	12.10
Share of suspected cancer patient treated within 62 Days (2015)	13,136	84.02	7.554	33.30	100
CV life satisfaction APS well-being data (2015)	13,136	0.989	0.394	0.570	3.050
Manufacturing employment share (2001)	13,136	0.153	0.0518	0.0538	0.337
Owned (outright + mortgage) share (2001)	13,136	0.684	0.0975	0.274	0.882

Notes: The table reports the number of observations (N), their mean, standard deviation (sd) as well as the minimum and maximum values. The summary statistics for the aggregate variables are reported based on the raw data, whereas in the regression tables these variables are used in standardized form.

Figure 1: Success rates by variable groupings



Notes: The graph plots success rates for different variable groupings. Light blue refers to models for our main sample with covariates comparable at the individual and aggregate levels. Orange relates to variables which are only available for smaller subsamples (individuals answering questions on housing or employment). Dark blue applies to models combining all available covariates in the main sample. Finally, green relates to individual-level variables which do not have a comparable grouping in the aggregate data.

Figure 2: Error rates and decomposition into false positives versus false negatives

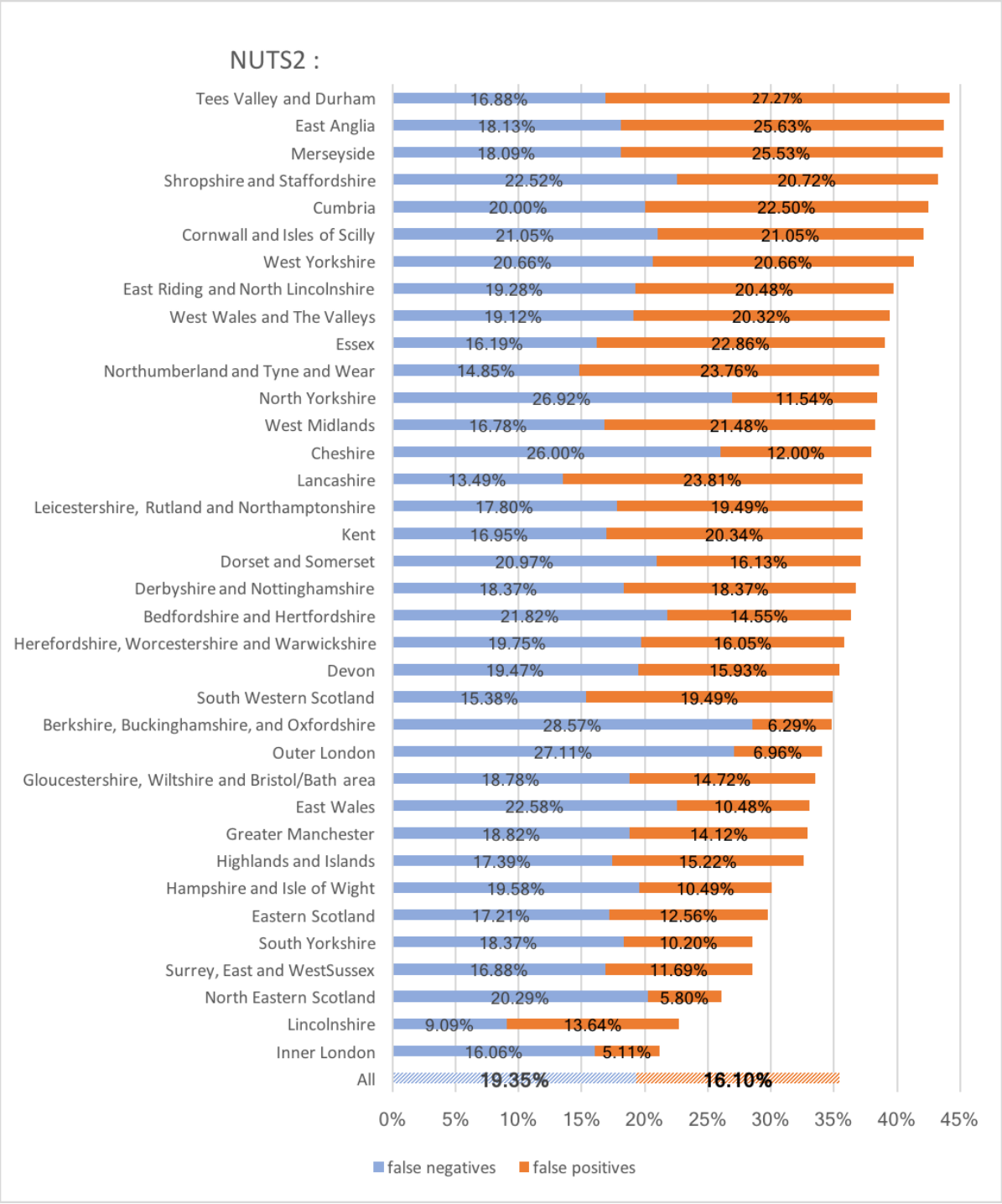
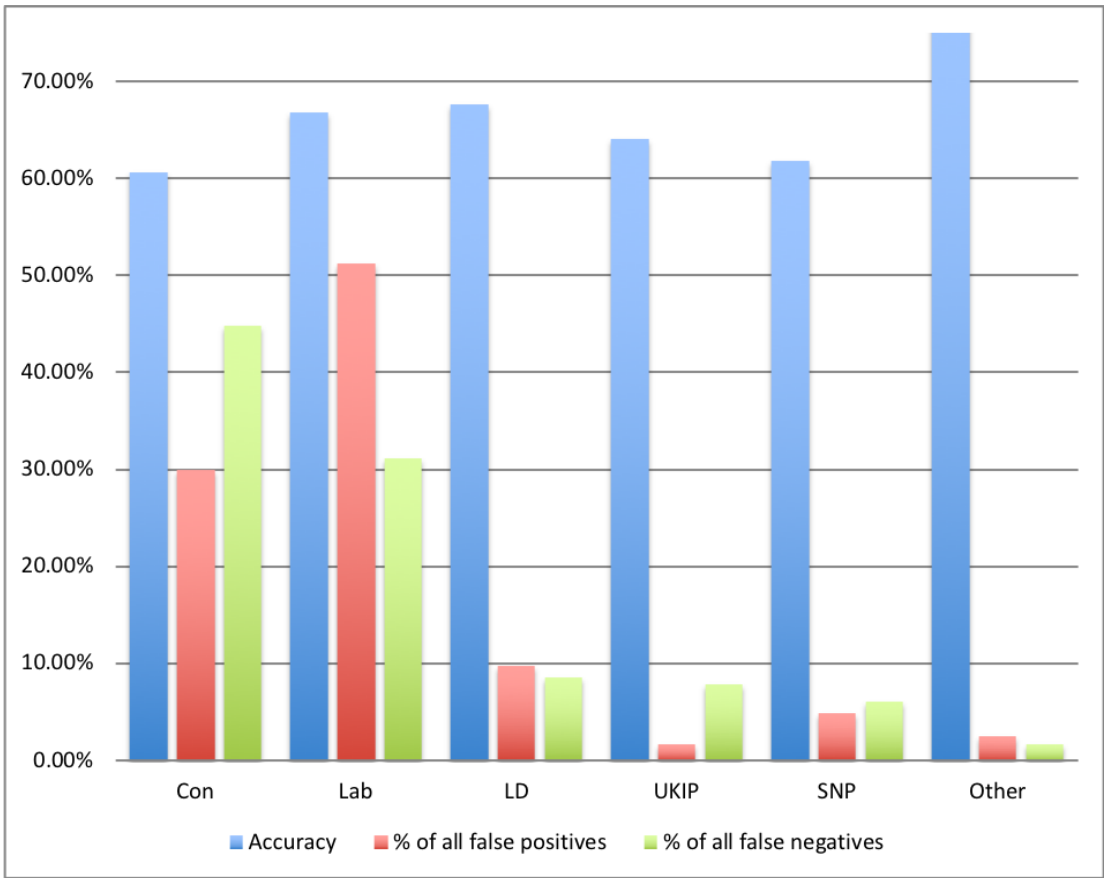


Figure 3: Overall accuracy and error decomposition by stated party preference (Conservative, Labour, Liberal Democrats, UKIP, SNP and Others)



Appendix to "Who Voted for Brexit? Individual and Regional Data Combined"

Eleonora Alabrese Sascha O. Becker Thiemo Fetzer Dennis Novy

August 2018

A Data and regression results

In this appendix we present our data and empirical regression results in more detail.

A.1 Sampling design

Concerning the design and data collection of *Understanding Society*, the general population sample is a stratified, clustered, equal probability sample of residential addresses drawn to a uniform design throughout the whole of the UK. For each wave, the data collection is spread over a two-year period, and the overall sample is divided into 24 monthly subsamples, each independently representative of the UK population. Computer assisted personal interviewing (CAPI) was mainly used to collect the data.[1]

A.2 Constructing the sample

The construction of our sample takes place in various steps. Initially, the raw individual survey (wave 8) consists of 21,076 observations. Then, matching the household survey leaves 20,821 individuals. Further matching with local authority codes results in a sample of 17,697 respondents (i.e. over 3000 surveyed individuals get lost because there is no location code associated with their households). Finally, we merge this last sample with the aggregate information used in Becker et al. (2017). In this last step, the number of surveyed individuals is 15,844 across 377 local authorities.

[1] These details are taken from Understanding Society: Design Overview by Buck and McFall (2012). For further details refer to the Understanding Society User Guide (wave 1-6) by Knies (2016).

When we consider the initial sample with 21,076 observations, 91% of the individuals provide an answer to the question concerning British EU membership. Among them, the share of those supporting Leave is 35.8%. Of the selected subsample with 15,844 units, 91.4% (14,476 individuals) disclose an answer for the outcome variable, and 42.6% turn out to be Leave supporters.[2]

As a final remark, we want to stress that our estimates come from the analysis of three specific subsamples of the 14,476 selected respondents. The main one contains 13,136 individuals. The sample with housing tenure status contains 6,425 individuals. The subsample on employment characteristics counts 8,434 individuals.

A.3 Regression results

We divide our variables into groupings as follows. The first group of explanatory variables includes basic demographic features such as sex, age, marital status, education and employment. The second group explores data on individuals' use of health services. The third group captures information on housing (ownership vs. renting) drawn from the household questionnaire. The fourth group refers to employment. This is followed by a focus on unearned income and state benefits. The sixth group consists of life satisfaction indicators. The seventh and final group covers nationality and ethnicity.

The results are reported in Tables A.1a to A.10. We present linear probability models as the default, with the exception of logit models in Table A.1b, probit models in Table A.1c and weighted OLS models in Table A.1d and A.1e.[3]

[2]In unreported tables (available upon request), we compare the 14,476 individuals individuals who answer the Brexit question to the 1368 non-respondents for each group of covariates (i.e. all regressors in Tables A.1a to A.10 in the appendix) and establish along which dimension the two groups are statistically different. If anything, non-respondents seem to display most of the characteristics of a typical Leave voter. More specifically, non-respondents are significantly older, less used to technology, with lower educational attainments and more frequently unemployed. In addition, they seek more medical attention, their housing status is more often local authority renting, and more of them receive income support. Finally, non-respondents are less often UK natives and more often members of an ethnic minority.

[3]We would like to note that sampling weights in Understanding Society, which we use inTable A.1d are quite homogenous. In our main estimation sample, the median sampling weight is 0.956, the 25th percentile is 0.770 and the 75th percentile is 1.237. This explains why weighted and unweighted regression results are so similar. [USOC wave 8 data has a substantive number of observations with missing weights. This is due to the fact that it is a pre-release version. The final version of wave 8 is expected to be released towards the end of 2018 or in

When variables are perfectly comparable at individual and aggregate levels, the first three columns of the tables directly compare those to address the potential ecological fallacy concern.

A.4 Demographics, technology, education and employment

In Tables A.1a to A.3 we present results from regressions based on different types of demographic characteristics. Tables A.1a to A.1e explore the relationship of voting Leave with sex, age and technology use. Table A.1a presents our baseline results estimated with a linear probability model (OLS). Tables A.1b and A.1c use the same explanatory variables but estimated with logistic and probit regressions, respectively, where we report marginal effects. Table A.1d reports weighted OLS regressions, with weights provided by *Understanding Society*. Table A.1e also displays weighted OLS regressions, but here we use artificial weights such that the proportion of Leave supporters in the sample matches the actual Brexit vote share. Overall, the coefficient signs and magnitudes are very similar across Tables A.1a to A.1c. They are also similar in comparison to Table A.1d and A.1e despite the weights and the reduced number of observations. We therefore focus our below discussion on Table A.1a.

Columns 1 to 3 of Table A.1a exhibit positive and significant coefficients for the old-age variables at both individual and aggregate levels, showing no evidence of ecological fallacy. Although the coefficient for the aggregate share of the elderly population is lower in magnitude, it presents a predictive power very similar to the individual counterpart. Column 4 indicates that males are 4.7% more likely to vote Leave. Compared to middle-aged respondents, the tendency to support Leave is substantially lower by 12.3% for younger cohorts up to the age of 30 and notably higher by 9.1% for individuals aged 60 or above. Columns 5 and 7 confirm these results in terms of significance even when we control for the share of the population aged 60 or above at the local authority level. In column 6 we focus on technology use. Individuals who do not use a smartphone are substantially more likely to vote Leave. Using the internet every day is associated with a substantially lower probability to vote Leave.

early 2019.] In Table A.1e, we mechanically re-weight the sample to align the share of 'Leave' voters with the actual Referendum result.

These patterns persist even once we control for sex and age in column 7.

In Table A.2 we explore the predictive power of educational attainment. Again, variables on education attainment relate to the referendum outcome in the same way and with matching power at both individual and aggregate levels although aggregate coefficients have lower magnitude and significance. Hence, highly qualified individuals with university and college degrees are considerably less likely to vote Leave by over 20% compared to people with average qualifications. In contrast, having no qualification is a very strong predictor of voting Leave. These results holds up once we control for aggregate characteristics on educational attainment in columns 3 and 5 as well as sex and age in column 6.

Next, in Table A.3 we analyze individuals' current employment and marital status. At the individual level, comparison groups are predominantly retired and divorced respondents, respectively.[4] Here, aggregate rates on employment are indistinguishable from zero (although they have the same predictive power as the individual variables, and self-employment and unemployment coefficients have the 'correct' sign). Column 1 of Table A.3 shows that self-employed and paid employees are more likely to support Remain (relative to mostly retired people). Column 4 shows that single and married people are significantly less likely to vote Leave (compared to divorcees, separated and widowed people). Again, most of these results hold up once we control for aggregate rates in column 3 as well as for age in column 5. Unemployment now also shows up as highly significant.[5]

To sum up our results on demographic variables, we find that individuals are more likely to support Leave if they are male, older, use less technology,

[4]Excluded categories among current activity feature Retired (64.7%), Looking after family or home (10%), Full-time student (14.3%), Long-term sick or disabled (7.5%), Doing something else (2.2%). Excluded categories among marital status feature Divorced (57.4%), Separated (10.3%), Widowed (31.6%), Other (0.7%).

[5]To get a sense of whether *changes* in (un)employment status matter, in unreported regressions, we used additional information based on a short employment history (looking at respondents participating in both wave 7 and the pre-release version of wave 8 with the EU question). The results suggest that the preferences for Remain and Leave are quite static or do not respond in a remarkable fashion to individuals switching employment status (by becoming unemployed or employed between wave 7 and wave 8). Rather, the first-order differences in tendencies to support Leave or Remain for our prediction exercise are driven by individuals who are employed or unemployed in both survey waves, implying that looking at only the cross-section is sufficient to capture the role of employment variables.

are less qualified, retired or unemployed, and divorced, separated or widowed. These findings are consistent with the results by Becker et al. (2017) based on aggregate data who also find that age, low educational attainment and unemployment are key explanatory variables to predict the Leave vote shares across UK voting areas.

A.5 Health

Table A.4 analyzes the relationship between Brexit support and individuals' use of health services. Interestingly, columns 1 and 2 show that individuals who visit their general practitioner (GP) very frequently (over ten times in the previous 12 months) are more likely to support Leave. Those are arguably individuals of poor health or older generations. Conversely, those who did not visited the GP even once have a slightly higher probability to support Remain. Controlling for age in column 2 turns the latter result insignificant (possibly because it is young people who do not go to the doctor) but preserves the former result on frequent GP visits.

A similar picture emerges from columns 3 and 4, focusing on individuals who are never or extremely often classified as out-patients. The same holds for people admitted as in-patients at least once during the preceding 12 months. That is, people of poor health as proxied by frequent visits to the GP or hospital are substantially more likely to support Leave. Perhaps it is therefore no coincidence that a key pledge of the pro-Brexit referendum campaign was to invest more in the National Health Service (NHS).

A.6 Housing

Table A.5 explores the role of property values for home owners and housing tenure (owned vs. rented). We note that due to many missing values, we only have 6,425 observations in this table.

When directly comparing individual tenure status to corresponding aggregate shares we see similar paths (columns 1 to 3), in particular with respect to direct ownership which is positively related to Leave support.

In terms of individual housing tenure, owning their own property tends

to make individuals more likely to support Leave, although this particular association is barely statistically significant. The omitted category here is renting through a housing association. More importantly, higher property values are significantly related to an increased likelihood of supporting Remain. A one-standard deviation increase in property values increases the Remain likelihood by roughly 4%. Property values are arguably positively linked to individuals' financial status, which would be consistent with earlier evidence on income based on aggregate data (see Becker et al., 2017).

A.7 Employment

This section shifts the focus towards employment-related determinants. For starters, Table A.6 indicates a higher probability of almost 10% to support Leave for individuals who did not work in the week prior to the questionnaire and who did not have a paid job compared to those respondents who were either working or had a paid job (stable across all specifications).

In Table A.7 we narrow our analysis to only those participants who worked or had a paid job. This reduces the number of observations to 8,434. First, columns 1 to 3 compare the individual sector of employment to the respective aggregate controls (manufacturing, construction, retail and finance as used in Becker et al., 2017). Estimates as well as their predictive power are aligned (although aggregate coefficients are lower in magnitude). Indeed, both specifications suggest that workers in the manufacturing, construction and retail industries are significantly more likely to support Leave. Note that individual estimates are fairly stable across all specifications.

In addition, it emerges from column 4 that those with a permanent job compared to those in non-permanent employment have a higher probability of supporting Leave. This result continues to hold qualitatively in column 5 after we control for individuals' age, sex and education as well as the sectoral distribution and growth of employment at the aggregate level in column 6. This result appears surprising, but we note that the subsample in Table A.7 is highly unbalanced in the sense that 90% of the respondents have a permanent job. Still, 60% of individuals with permanent jobs support Remain versus 70% of those

with temporary jobs. It also appears likely that the very young respondents, who are overwhelmingly in favor of Remain, are less likely to hold permanent jobs. Our age dummies in column 5 might not pick up these age patterns appropriately. Finally, self-employed respondents are also more likely to support Leave, even though this association is insignificant for most specifications in the table.

Overall, consistent with the aggregate results in Becker et al. (2017) our findings support the view that individuals are more willing to vote for Brexit if they work in sectors such as manufacturing that have arguably been hit relatively hard by trade openness and international competition (also see Colantone and Stanig, 2017). In addition, workers in manufacturing, construction and retail sectors have lower educational attainment on average while the opposite is true for workers in financial sector.

A.8 Unearned income and state benefits

In Table A.8 we highlight the role of unearned income and state benefits. In column 1 we find that respondents who receive core benefits have significantly raised probability of supporting Leave compared to those receiving none. These core benefits are broken down into their various components in column 2. In particular, recipients of income support are substantially more likely to be in favor of Leave (by 20%), whereas job seeker's allowance, child benefit and universal credit do not matter.

Similar results hold for people receiving pensions. This particular finding is likely driven by the overwhelming share of older people amongst pension receivers (see section A.4). The same pattern holds for people on disability benefits, in line with our estimates on health service usage (see section A.5).

Finally, the opposite is true for respondents who receive other sources of income. Those are broken down in column 3. The key income streams are education grants and student loans as well as payments from family members living elsewhere. This suggests a tight link previously with age and education (see section A.4).

In summary, the forms of income and benefits in Table A.8 are likely cor-

related with more fundamental characteristics such as age and health, as discussed in previous tables.

A.9 Life satisfaction

In Table A.9 we explore the potential link between Brexit support and indices of health, income and life satisfaction. When looking at overall life satisfaction only (columns 1 to 3), the individual coefficients suggest that dissatisfied people are significantly more likely to favor Leave while the aggregate estimate implies 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 Leave support. Success rates of prediction are very similar whichever level of variation is considered.

In addition, people dissatisfied with health and income have a higher probability of supporting Leave by 5.5% and 6.4%, respectively. Once again, we can relate these findings to those in Table A.4 on health and Table A.8 on income and benefits. Interestingly, people dissatisfied with their amount of leisure time are significantly more likely to support Remain by 6.3%. This may be linked to the fact that these respondents have on average higher levels of educational attainment and they are generally younger. Note that when these individual variables are considered (columns 4 and 5) the individual estimate of overall life satisfaction is absorbed and becomes insignificant.

A.10 Nationality and ethnicity

Table A.10 provides insights on the importance of individuals' nationality and ethnicity in shaping their attitudes towards Brexit. Survey participants born in the UK as opposed to elsewhere have a significantly larger probability of supporting Leave by 12.4% (see column 1). It is useful to point out that in the sample, 90% of respondents are born in the UK, and 95% of them are white.

In terms of ethnic minorities compared to whites (see column 2), people of mixed ethnicity, Asians and black respondents all have a significantly larger probability of supporting Remain (in the range of 12% to 23%). These results are in line with the preliminary work by Martin et al. (2017).

Finally, aggregate controls for migration are insignificant with the exception of the EU share of migrants in 2001, which is positive linked with support for Remain.

Appendix tables

Table A.1a: Demographics: Sex, Age and Technology Use (OLS)

	Should the UK leave the EU						
	Ecological fallacy						
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)	(6)	(7)
Sex = Male				0.0471***	0.0473***		0.0502***
				(0.0073)	(0.0073)		(0.0072)
Age = 30 or younger				-0.1226***	-0.1192***		-0.1092***
				(0.0130)	(0.0126)		(0.0128)
Age = 60 or older	0.1240***		0.1174***	0.0911***	0.2808***		0.2367***
	(0.0113)		(0.0112)	(0.0116)	(0.0729)		(0.0726)
Share population 60 or older (2001)		0.0339***	0.0280***		0.0365***		0.0364***
		(0.0069)	(0.0069)		(0.0076)		(0.0075)
Age = 60 or older * Share population 60 or older (2001)					-0.0925***		-0.0917***
					(0.0336)		(0.0332)
Use smartphone = No						0.0776***	0.0382***
						(0.0126)	(0.0129)
Has mobile computing device = No						0.0134	0.0214**
						(0.0099)	(0.0098)
Frequency using internet = Every day						-0.1018***	-0.0760***
						(0.0148)	(0.0148)
Frequency using internet = No access						0.0358	0.0320
						(0.0398)	(0.0400)
Constant	0.3845***	0.4221***	0.3864***	0.3954***	0.3363***	0.4817***	0.3907***
	(0.0093)	(0.0083)	(0.0088)	(0.0097)	(0.0237)	(0.0154)	(0.0261)
Observations	13,136	13,136	13,136	13,136	13,136	13,136	13,136
Predictive success rate (from logit)	0.5820	0.5827	0.5834	0.5891	0.5910	0.5887	0.5917

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.1b: Demographics: Sex, Age and Technology Use (Logit)

	Should the UK leave the EU						
	Ecological fallacy						
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)	(6)	(7)
Sex = Male				0.0480***	0.0484***		0.0518***
				(0.0074)	(0.0074)		(0.0074)
Age = 30 or younger				-0.1265***	-0.1238***		-0.1147***
				(0.0134)	(0.0131)		(0.0135)
Age = 60 or older	0.1240***		0.1176***	0.0904***	0.2899***		0.2489***
	(0.0113)		(0.0112)	(0.0115)	(0.0696)		(0.0716)
Share population 60 or older (2001)		0.0340***	0.0284***		0.0382***		0.0382***
		(0.0070)	(0.0071)		(0.0081)		(0.0081)
Age = 60 or older * Share population 60 or older (2001)					-0.0980***		-0.0975***
					(0.0338)		(0.0338)
Use smartphone = No						0.0779***	0.0379***
						(0.0127)	(0.0129)
Has mobile computing device = No						0.0136	0.0221**
						(0.0100)	(0.0101)
Frequency using internet = Every day						-0.1019***	-0.0761***
						(0.0148)	(0.0149)
Frequency using internet = No access						0.0358	0.0319
						(0.0409)	(0.0412)
Observations	13,136	13,136	13,136	13,136	13,136	13,136	13,136

Notes: The table reports results from Logit regressions in terms of marginal effects. Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.1c: Demographics: Sex, Age and Technology Use (Probit)

	Should the UK leave the EU						
	Ecological fallacy						
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)	(6)	(7)
Sex = Male				0.0478***	0.0481***		0.0513***
				(0.0074)	(0.0074)		(0.0074)
Age = 30 or younger				-0.1257***	-0.1233***		-0.1139***
				(0.0133)	(0.0131)		(0.0134)
Age = 60 or older	0.1240***		0.1176***	0.0906***	0.2903***		0.2489***
	(0.0113)		(0.0112)	(0.0116)	(0.0704)		(0.0721)
Share population 60 or older (2001)		0.0341***	0.0285***		0.0382***		0.0383***
		(0.0070)	(0.0071)		(0.0081)		(0.0080)
Age = 60 or older * Share population 60 or older (2001)					-0.0980***		-0.0975***
					(0.0339)		(0.0338)
Use smartphone = No						0.0778***	0.0379***
						(0.0126)	(0.0129)
Has mobile computing device = No						0.0135	0.0219**
						(0.0100)	(0.0101)
Frequency using internet = Every day						-0.1018***	-0.0761***
						(0.0148)	(0.0148)
Frequency using internet = No access						0.0358	0.0321
						(0.0406)	(0.0409)
Observations	13,136	13,136	13,136	13,136	13,136	13,136	13,136

Notes: The table reports results from Probit regressions in terms of marginal effects. Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.1d: Demographics: Sex, Age and Technology Use (Weighted OLS using USOC sampling weights)

	Should the UK leave the EU						
	Ecological fallacy						
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)	(6)	(7)
Sex = Male				0.0359***	0.0359***		0.0390***
				(0.0120)	(0.0120)		(0.0120)
Age = 30 or younger				-0.1265***	-0.1255***		-0.1149***
				(0.0203)	(0.0204)		(0.0201)
Age = 60 or older	0.1393***		0.1351***	0.1030***	0.2986***		0.2608***
	(0.0133)		(0.0131)	(0.0143)	(0.0893)		(0.0903)
Share population 60 or older (2001)		0.0243***	0.0175**		0.0272***		0.0266***
		(0.0085)	(0.0084)		(0.0100)		(0.0100)
Age = 60 or older * Share population 60 or older (2001)					-0.0943**		-0.0951**
					(0.0413)		(0.0414)
Use smartphone = No						0.0818***	0.0355**
						(0.0167)	(0.0178)
Has mobile computing device = No						0.0004	0.0095
						(0.0147)	(0.0146)
Frequency using internet = Every day						-0.1089***	-0.0737***
						(0.0190)	(0.0189)
Frequency using internet = No access						-0.0021	-0.0074
						(0.0464)	(0.0468)
Constant	0.3901***	0.4332***	0.3910***	0.4093***	0.3488***	0.5019***	0.4040***
	(0.0108)	(0.0092)	(0.0106)	(0.0123)	(0.0310)	(0.0201)	(0.0360)
Observations	8,188	8,188	8,188	8,188	8,188	8,188	8,188

Notes: The table reports results from weighted linear probability regressions, using *Understanding Society* sampling weights (weighted OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.1e: Demographics: Sex, Age and Technology Use (Weighted OLS using weights that mechanically align share of 'Leave' voters with Referendum result)

	Should the UK leave the EU						
	Ecological fallacy						
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)	(6)	(7)
Sex = Male				0.0480***	0.0481***		0.0510***
				(0.0074)	(0.0074)		(0.0073)
Age = 30 or younger				-0.1288***	-0.1255***		-0.1152***
				(0.0138)	(0.0134)		(0.0137)
Age = 60 or older	0.1254***		0.1188***	0.0916***	0.2916***		0.2478***
	(0.0113)		(0.0112)	(0.0116)	(0.0733)		(0.0731)
Share population 60 or older (2001)		0.0347***	0.0288***		0.0379***		0.0379***
		(0.0071)	(0.0071)		(0.0079)		(0.0078)
Age = 60 or older * Share population 60 or older (2001)					-0.0974***		-0.0966***
					(0.0337)		(0.0334)
Use smartphone = No						0.0782***	0.0380***
						(0.0126)	(0.0128)
Has mobile computing device = No						0.0136	0.0218**
						(0.0101)	(0.0100)
Frequency using internet = Every day						-0.1023***	-0.0760***
						(0.0147)	(0.0146)
Frequency using internet = No access						0.0332	0.0295
						(0.0380)	(0.0382)
Constant	0.4343***	0.4730***	0.4362***	0.4456***	0.3839***	0.5320***	0.4378***
	(0.0096)	(0.0085)	(0.0092)	(0.0100)	(0.0239)	(0.0153)	(0.0261)
Observations	13,136	13,136	13,136	13,136	13,136	13,136	13,136

Notes: The table reports results from weighted linear probability regressions (weighted OLS). Weights mechanically reproduce the Brexit referendum result. Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: Demographics: Education

	Should the UK leave the EU					
	Ecological fallacy					
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)	(6)
Highest qualification = Degree	-0.2590***		-0.2407***	-0.2348***	-0.2162***	-0.2394***
	(0.0115)		(0.0119)	(0.0119)	(0.0123)	(0.0123)
Highest qualification = Other higher degree	-0.0842***		-0.0793***	-0.0601***	-0.0556***	-0.0796***
	(0.0155)		(0.0156)	(0.0157)	(0.0157)	(0.0161)
Highest qualification = Other lower qualification				0.1478***	0.1473***	0.1019***
				(0.0177)	(0.0176)	(0.0171)
Highest qualification = No qualification	0.0988***		0.0948***	0.1229***	0.1190***	0.0822***
	(0.0181)		(0.0178)	(0.0186)	(0.0183)	(0.0184)
Share of res. pop. qualification 4+ (2001)		-0.0442***	-0.0292**		-0.0220*	-0.0171
		(0.0134)	(0.0128)		(0.0131)	(0.0127)
Share of res. pop. no qualifications (2001)		0.0244*	0.0180		0.0210*	0.0285**
		(0.0132)	(0.0127)		(0.0123)	(0.0119)
Share population 60 or older (2001)					0.0145**	0.0086
					(0.0065)	(0.0064)
Sex = Male						0.0520***
						(0.0072)
Age = 30 or younger						-0.1637***
						(0.0123)
Age = 60 or older						0.0323***
						(0.0119)
Constant	0.4968***	0.4223***	0.4915***	0.4727***	0.4673***	0.4798***
	(0.0098)	(0.0079)	(0.0092)	(0.0104)	(0.0096)	(0.0105)
Observations	13,136	13,136	13,136	13,136	13,136	13,136
Predictive success rate (from logit)	0.5940	0.5887	0.6052	0.6092	0.6135	0.6404

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Demographics: Employment and Marital Status

	Should the UK leave the EU				
	Ecological fallacy				
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)
Current activity = Self-employed	-0.0489***		-0.0473**		-0.0263
	(0.0184)		(0.0183)		(0.0195)
Current activity = In paid employment	-0.0773***		-0.0775***		-0.0336***
	(0.0112)		(0.0112)		(0.0125)
Current activity = Unemployed	0.0199		0.0161		0.0896***
	(0.0266)		(0.0269)		(0.0269)
Self-employment rate (2015)		-0.0026	-0.0043		-0.0052
		(0.0082)	(0.0081)		(0.0079)
Employment rate (2015)		0.0027	0.0049		0.0027
		(0.0113)	(0.0113)		(0.0111)
Unemployment rate (2015)		0.0156	0.0157		0.0188*
		(0.0114)	(0.0114)		(0.0112)
Current legal marital status = Single				-0.1714***	-0.0775***
				(0.0149)	(0.0166)
Current legal marital status = Married or civil partner				-0.0890***	-0.0679***
				(0.0136)	(0.0139)
Age = 30 or younger					-0.1213***
					(0.0160)
Age = 60 or older					0.0639***
					(0.0136)
Constant	0.4653***	0.4223***	0.4654***	0.5201***	0.5006***
	(0.0107)	(0.0087)	(0.0108)	(0.0137)	(0.0175)
Observations	13,136	13,136	13,136	13,136	13,136
Predictive success rate (from logit)	0.5839	0.5839	0.5843	0.5937	0.5921

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Health

	Should the UK leave the EU					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Visits GP in 12m = None	-0.0271**	-0.0113				
	(0.0107)	(0.0106)				
Visits GP in 12m = Over 10	0.0892***	0.0847***				
	(0.0203)	(0.0199)				
Age = 60 or older		0.1213***		0.1200***		0.1226***
		(0.0113)		(0.0114)		(0.0114)
Share of suspected cancer patient treated within 62 Days (2015)		-0.0213**		-0.0211**		-0.0214**
		(0.0088)		(0.0087)		(0.0088)
Out-patient in 12m = None			-0.0331***	-0.0166*		
			(0.0085)	(0.0086)		
Out-patient in 12m = Over 10			0.0791***	0.0717***		
			(0.0277)	(0.0271)		
In-patient in 12m = Yes					0.0475***	0.0353**
					(0.0150)	(0.0150)
Constant	0.4226***	0.3826***	0.4379***	0.3926***	0.4182***	0.3820***
	(0.0091)	(0.0096)	(0.0097)	(0.0103)	(0.0088)	(0.0091)
Observations	13,136	13,136	13,136	13,136	13,136	13,136
Predictive success rate (from logit)	0.5818	0.5974	0.5880	0.5910	0.5839	0.5944

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Housing

	Should the UK leave the EU					
	Ecological fallacy					
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)	(6)
Value of property: home owners				-0.0428***	-0.0422***	-0.0408***
				(0.0111)	(0.0114)	(0.0112)
Housing tenure = Owned (outright + mortgage)	0.1595**		0.1269*	0.1141*	0.1143*	0.1303*
	(0.0760)		(0.0694)	(0.0626)	(0.0682)	(0.0712)
Housing tenure = Rented from employer or private	-0.0109		0.0100		-0.0020	0.0206
	(0.1831)		(0.1681)		(0.1675)	(0.1688)
Housing tenure = Local authority rent	0.2391		0.1811			0.2285
	(0.1923)		(0.1880)			(0.1836)
Age = 60 or older				0.1311***	0.1317***	0.1318***
				(0.0151)	(0.0151)	(0.0151)
Owned (outright + mortgage) share (2001)		0.0799***	0.0783***	0.0372***	0.0201	0.0469**
		(0.0183)	(0.0184)	(0.0069)	(0.0180)	(0.0214)
Owned (outright + mortgage) share growth (2001-2011)				-0.0184**	-0.0391*	-0.0613***
				(0.0075)	(0.0214)	(0.0232)
Private rented share (2001)		0.0132	0.0132		-0.0139	0.0013
		(0.0116)	(0.0116)		(0.0119)	(0.0134)
Privated rented share growth (2001-2011)					-0.0170	-0.0396*
					(0.0209)	(0.0228)
Council rented share (2001)		0.0360**	0.0350**			0.0380**
		(0.0174)	(0.0174)			(0.0189)
Council rented share growth (2001-2011)						-0.0115
						(0.0093)
Constant	0.2609***	0.4187***	0.2928***	0.2611***	0.2607***	0.2444***
	(0.0757)	(0.0076)	(0.0691)	(0.0622)	(0.0678)	(0.0708)
Observations	6,425	6,425	6,425	6,425	6,425	6,425
Predictive success rate (from logit)	0.5772	0.5729	0.5729	0.6026	0.6002	0.6050

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Two-way clustered standard errors at the local authority and household levels are presented in parentheses, asterisks indicate *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.6: Current Employment: All Individuals

	Should the UK leave the EU			
	Ecological fallacy			
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)
No work last week & doesn't have paid job	0.0979***		0.0970***	0.0939***
	(0.0101)		(0.0101)	(0.0098)
Unemployment rate (2015)		0.0144*	0.0127	0.0171**
		(0.0083)	(0.0082)	(0.0074)
Manufacturing employment share (2001)				0.0347***
				(0.0079)
Manufacturing employment share change (2001-2011)				0.0042
				(0.0081)
Construction employment share (2001)				0.0086
				(0.0105)
Construction employment share change (2001-2011)				0.0088
				(0.0098)
Retail employment share (2001)				0.0361***
				(0.0078)
Retail employment share change (2001-2011)				-0.0237***
				(0.0068)
Finance employment share (2001)				0.0023
				(0.0078)
Finance employment share change (2001-2011)				-0.0131
				(0.0085)
Constant	0.3852***	0.4223***	0.3856***	0.3866***
	(0.0097)	(0.0087)	(0.0098)	(0.0084)
Observations	13,136	13,136	13,136	13,136
Predictive success rate (from logit)	0.5839	0.5839	0.5797	0.5942

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Current Employment: Individuals With Paid Jobs

	Should the UK leave the EU					
	Ecological fallacy					
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)	(6)
Current job = Permanent				0.0924***	0.0580***	0.0504***
				(0.0177)	(0.0176)	(0.0174)
Current job sector = Manufacturing	0.1395***		0.1159***	0.1352***	0.0731***	0.0597***
	(0.0191)		(0.0187)	(0.0191)	(0.0191)	(0.0186)
Current job sector = Construction	0.1704***		0.1604***	0.1659***	0.0858***	0.0811***
	(0.0268)		(0.0266)	(0.0269)	(0.0260)	(0.0257)
Current job sector = Wholesale & retail	0.0877***		0.0747***	0.0854***	0.0350**	0.0284*
	(0.0177)		(0.0176)	(0.0177)	(0.0172)	(0.0172)
Current job sector = Finance	-0.0449		-0.0415	-0.0484*	-0.0437	-0.0426
	(0.0277)		(0.0274)	(0.0278)	(0.0275)	(0.0272)
Current job = Self-employed	0.0267		0.0301*	0.0380**	0.0136	0.0153
	(0.0177)		(0.0173)	(0.0175)	(0.0168)	(0.0166)
Sex = Male					0.0328***	0.0329***
					(0.0099)	(0.0100)
Age = 30 or younger					-0.1338***	-0.1306***
					(0.0142)	(0.0138)
Age = 60 or older					0.0184	0.0200
					(0.0189)	(0.0185)
Highest qualification = Degree					-0.2714***	-0.2563***
					(0.0124)	(0.0127)
Highest qualification = Other higher degree					-0.1020***	-0.0979***
					(0.0185)	(0.0188)
Manufacturing employment share (2001)		0.0470***	0.0437***			0.0332***
		(0.0092)	(0.0091)			(0.0091)
Manufacturing employment share change (2001-2011)						0.0005
						(0.0082)
Construction employment share (2001)		0.0250**	0.0236**			0.0038
		(0.0109)	(0.0107)			(0.0098)
Construction employment share change (2001-2011)						-0.0001
						(0.0093)
Retail employment share (2001)		0.0386***	0.0377***			0.0260***
		(0.0088)	(0.0087)			(0.0083)
Retail employment share change (2001-2011)						-0.0143*
						(0.0077)
Finance employment share (2001)		0.0086	0.0106			0.0134
		(0.0092)	(0.0093)			(0.0090)
Finance employment share change (2001-2011)						-0.0178**
						(0.0090)
Self-employment rate (2015)		0.0149*	0.0144			0.0127
		(0.0089)	(0.0088)			(0.0089)
Unemployment rate (2015)						0.0161**
						(0.0078)
Constant	0.3550***	0.3874***	0.3583***	0.2709***	0.4354***	0.4375***
	(0.0103)	(0.0084)	(0.0095)	(0.0195)	(0.0210)	(0.0200)
Observations	8,434	8,434	8,434	8,434	8,434	8,434
Predictive success rate (from logit)	0.6111	0.6204	0.6132	0.6118	0.6370	0.6560

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Unearned Income and State Benefits

	Should the UK leave the EU		
VARIABLES	(1)	(2)	(3)
Receives core benefits	0.0357***		
	(0.0124)		
Receives pensions	0.1177***		
	(0.0115)		
Receives disability benefits	0.1119***		
	(0.0167)		
Receives other benefits or credits	0.0867***		
	(0.0131)		
Receives other sources of income	-0.0496***		
	(0.0152)		
Core benefits:			
Income Support		0.2002***	
		(0.0354)	
Job Seeker's Allowance		0.0467	
		(0.0404)	
Child Benefit		0.0006	
		(0.0131)	
Universal Credit		0.0357	
		(0.0472)	
Other sources of income:			
Education Grant other than a Student Loan or Tuition Fee Loan			-0.2329***
			(0.0286)
Trade Union or Friendly Society Payment			0.1427
			(0.1883)
Maintenance or Alimony			0.0636*
			(0.0361)
Payments from a family member not living with you			-0.1063***
			(0.0402)
Rent from Boarders or Lodgers (not family members) living here with you			-0.0823
			(0.0646)
Rent from any other property even if that only covers that property's mortg			-0.0317
			(0.0201)
Or any other regular payment			-0.0812**
			(0.0381)
Constant	0.3615***	0.4182***	0.4287***
	(0.0102)	(0.0090)	(0.0089)
Observations	13,136	13,136	13,136
Predictive success rate (from logit)	0.5977	0.5875	0.5841

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.9: Life Satisfaction

	Should the UK leave the EU				
	Ecological fallacy				
VARIABLES	(1) Individual	(2) Aggregate	(3) Both	(4)	(5)
Dissatisfied with health				0.0549***	0.0533***
				(0.0116)	(0.0116)
Dissatisfied with income				0.0643***	0.0652***
				(0.0121)	(0.0121)
Dissatisfied with amount of leisure time				-0.0625***	-0.0629***
				(0.0111)	(0.0111)
Dissatisfied with life overall	0.0252**		0.0262**	-0.0101	-0.0085
	(0.0125)		(0.0125)	(0.0156)	(0.0156)
CV life satisfaction APS well-being data (2015)		0.0263***	0.0265***		0.0263***
		(0.0073)	(0.0073)		(0.0073)
Constant	0.4187***	0.4223***	0.4186***	0.4107***	0.4108***
	(0.0089)	(0.0085)	(0.0087)	(0.0095)	(0.0093)
Observations	13,136	13,136	13,136	13,136	13,136
Predictive success rate (from logit)	0.5839	0.5848	0.5866	0.5804	0.5850

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.

Table A.10: Nationality and Ethnicity

	Should the UK leave the EU			
VARIABLES	(1)	(2)	(3)	(4)
Born in UK	0.1239***		0.0706***	0.0581***
	(0.0160)		(0.0184)	(0.0182)
Ethnic group = Mixed		-0.2298***	-0.2154***	-0.2020***
		(0.0274)	(0.0271)	(0.0280)
Ethnic group = Asian		-0.1186***	-0.0817***	-0.0823***
		(0.0228)	(0.0252)	(0.0266)
Ethnic group = Black		-0.1864***	-0.1497***	-0.1301***
		(0.0210)	(0.0222)	(0.0281)
Ethnic group = Other ethnic group		-0.0687	-0.0274	-0.0252
		(0.0549)	(0.0561)	(0.0594)
EU migrant resident share (2001)				-0.0382***
				(0.0112)
Non-EU migrant resident share (2001)				0.0150
				(0.0170)
EU migrant resident growth (2001-2011)				0.0083
				(0.0139)
Non-EU migrant resident growth (2001-2011)				-0.0136
				(0.0128)
Constant	0.3102***	0.4379***	0.3706***	0.3812***
	(0.0151)	(0.0092)	(0.0194)	(0.0195)
Observations	13,136	13,136	13,136	13,136
Predictive success rate (from logit)	0.5839	0.5839	0.5839	0.5850

Notes: The table reports results from linear probability regressions (OLS). Non-dummy variables are standardized. Authority-level clustered standard errors are presented in parentheses, asterisks indicate *** p<0.01, ** p<0.05, * p<0.1.