DISCUSSION PAPER SERIES

IZA DP No. 9570

Predicting the Irish "Gay Marriage" Referendum

Nikos Askitas

December 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Predicting the Irish "Gay Marriage" Referendum

Nikos Askitas
IZA

Discussion Paper No. 9570
December 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 9570
December 2015

ABSTRACT

Predicting the Irish "Gay Marriage" Referendum*

On February 20 2015 Irish Premier Enda Kenny confirmed that a "yes-no" referendum on same sex marriage would be held on May 22 of the same year. A yes vote would legalise same sex marriage in Ireland. As the Irish premier put it, the vote was about "tolerance, respect and sensitivity". The electoral outcome turned out to be 62.07% for the yes vote with voter turnout at 60.52% of the registered voters. Ireland thus became the first country in the world to legalise same sex marriage through a popular vote. Using hourly Google Search data one week prior to the Irish Referendum of May 22 2015 and a simple ratio of "vote yes" to "vote no" searches I demonstrate how the outcome could have been predicted on the nose. The method is used here successfully for the second time and is so far as I know the only one which forecasts popular vote with Google Search.

JEL Classification: D72, G34

Keywords: referendum, predicting, Google Trends, Google Search, complexity, behaviour, data science, computational social science, complex systems

Corresponding author:

Nikos Askitas
Institute for the Study of Labor (IZA)
Schaumburg-Lippe-Strasse 5-9
53113 Bonn
Germany
E-mail: askitas@iza.org

* I would like to thank the Google Trends Team at Google for their help on every question I confront them with.

1. INTRODUCTION

In Askitas (2015a) I was able to perform a live nowcasting of the Greek yes-no referendum held on June 27 2015. It's outcome was surprising and traditional polls were unable to pick up the outcome. I was able to call the outcome live and well ahead of the exit polls. I subsequently applied the same method to nowcast housing prices as well in Askitas (2015b) with good results. While in the case of the Greek referendum I exploited the yes-no dichotomy in the case of housing prices I used the dichotomy of buy and sell to nowcast prices of "repeat house sales". In this note I would like to do the same exercise as in Askitas (2015a) for the Irish Referendum held May 22 of 2015. The methodology is the same, it is fairly simple but novel and I know of no other application of this type. It allows us to once again get the outcome on the nose and demonstrates that the power in predicting does not so much come from the method but from the data.

The formation of opinions and convictions and well as that of consensus and majority are highly endogenous processes were agents are driven by their tendency to be near some notion of the weighted average of their peers on the one hand and their desire to not completely surrender their individualism on the other. Norms, institutions and culture are confronted with feelings and emotions and mixed with cognitive shortcoming of psychological biases, within and among individuals' conflicts and centre stage advocacy they create a highly dynamic and complex system which offers itself to a new brand of social science the computational one. While we cannot necessarily predict the future of such systems we are sometimes able to observe their evolution as it occurs and this is what Google Trends provides us with: a way to gauge socioeconomic process in real time.

In this note I rewind reality (another of the nice properties Google Trends as a data provisioning tool shares with other digital recording devices) and replay the last week before the Irish referendum on gay marriage seeking to provide further support for my claim that popular vote is predictable (in the sense of nowcasting)[1].

The reader should be reminded[2] that although three polls published on the weekend preceding Monday May 18 had the yes vote ahead, the situation was far from clear as the yes vote was weakening. A poll by Ipsos MRBI for the Irish Times had yes at 70%, Millward Brown for the Sunday Independent had it on 69% and a Red C poll for the Sunday Business Post had it 73% but taking into account the undecided the Millward Brown forecast was: yes at 53%, no at 24% and don?t know at 23% while in March, the figures were 66%, 21% and 13% respectively. The Ipsos figures were also recording a downwards trend for the yes vote. As we will see Google Trends shows us exactly were the rebound occurred for the the yes vote.

The rest of this note has a simple structure. In Section 2 I discuss Google Trends as a data provisioning tool while in Section 3 I discuss the methodology as well as the strengths and weaknesses of the identification strategy. In Section 4 I discuss the approach and the result of

[1] Of course forecasting is self destructive as it is well know. As soon as we start using Google-Trends to monitor voter sentiment the actors of the voting process may well Google search in a strategic manner in an attempt to influence the vote. In this case of course we would simply get better volumes and more robust data provided all sides engage in "strategic" Google searching with the same intensity.

[2] See www.theguardian.com.

the nowcasting and finally I discuss conclusions in Section 5.

2. DATA

Google Trends data[3] is relative data. Within an aggregation time unit i (which can be a hour, a day or a week) we take the number x_i of searches which include our keyword of interest $\mathtt{x}$ and divide that by the total number of searches T_i in the same aggregation time unit i, so that we form x_i/T_i. Moreover if we are observing a certain time period (which can be seven days for hourly data, three months for daily data and everything since 2004 for weekly data) then $i = 1 \dots n$ for some n ($n = 7 \times 24$ in case of hourly data or about 3×30 in case of daily or the number of weeks since 2004 in case of weekly data). If then $M_n = \max_{i=1\dots n}\{x_i/T_i\}$ then the time series we get from Google is:

$$G_i(\mathtt{x}) = \frac{100 \cdot x_i}{T_i \cdot M_n}$$

or setting $c_n = 100/M_n$

$$G_i(\mathtt{x}) = \frac{x_i}{T_i} c_n. \tag{2.1}$$

Google uses undisclosed, proprietary algorithms to classify and group searches into categories such as Travel, Real Estate, Business, Health etc. The final piece of Google trends nomenclature we need to explain in order to proceed with the description of the data is the exclusion mechanism. One can ask for all searches containing a certain keyword without searches which contain certain others; up to 30 keywords can be excluded. For examples drawing the time series for "$\mathtt{x} \quad -\mathtt{y}_1 \quad \cdots \quad -\mathtt{y}''_{30}$ will produce the relative volumes of all searches which contain the word $\mathtt{x}$ without those that contain any of $\mathtt{y}_1, \cdots, \mathtt{y}_{30}$.

In Askitas (2015a) I simply looked at searches containing `yes` and those containing `no` to successfully and precisely nowcast the Greek Referendum of July 5 2015. I used the exclusion mechanism to break down such searches and could see that all other keywords involved were in accordance with voter intent and hence such searches had to have been made by people seeking like-minded ones and respective rallies to join. The extreme polarisation and the fact that this was a singular societal event made that identification problem easy. For the Irish case it is somewhat different but not hopeless. I look at two kinds of searches:

"`vote yes`" and "`vote no`".

Assuming for a minute that these types of searches are faithful proxies of the respective voter intent we can get shares of yes and no intentions among those who are active on Google. The time series I look at are depicted in Figure 1.

[3]The description of the Google Trends data here draws from from the data section of Askitas (2015a) and is provided for completion and the convenience of the reader.

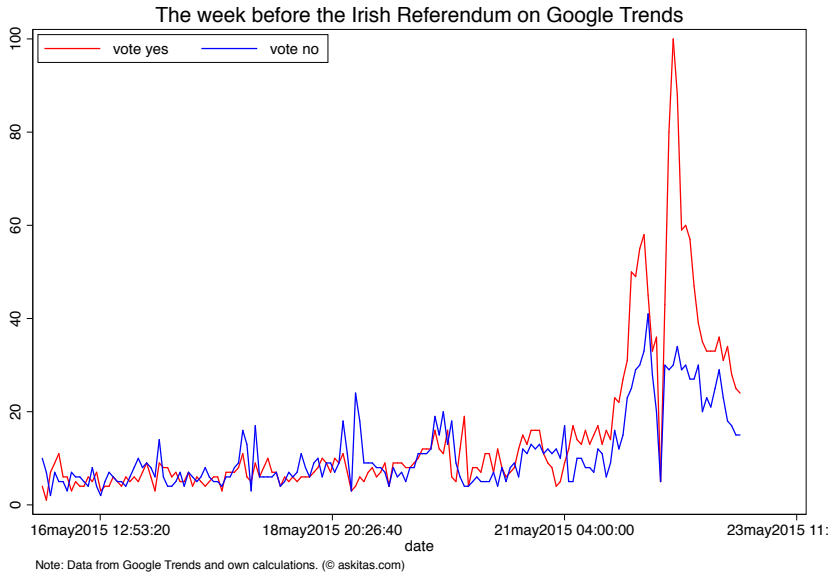

FIGURE 1.— The Irish referendum week on Google Trends. Hourly earches for "vote yes" and "vote no" are proxies for voter intent.

The hourly time series is published in 7-day segments. Data Source: Google Trends (www.google.com/trends).

Search intensities are vulnerable to ambient search noise and shocks from irrelevant keywords in other words from random variation of the denominator in equation (2.1) hence I will be looking at the "yes to no ratio" just like I did in Askitas (2015a). In other words the series that I will form is the point-wise ratios of the "vote yes" and "vote no" series and from that actual shares. This series has the advantage that it equals the ratio of the absolute number of yes searches to the absolute number of no searches in other words it is no longer vulnerable to the denominator of equation (2.1). The yes to no ratio series and its twelve week moving average are depicted in Figure 2.

3. METHODOLOY AND IDENTIFICATION

By taking the ratio y/x of the search intensities of two tokens (y=yes and x=no) I convert two relative Google search intensity ratios (for the two options of a dichotomy) to one ratio of the absolute number of searches of the two options. Since the space of decisions consists of these two option we can now get shares of the two options in the space of those agents that are active on Google by simply taking $y/(x+y)$. Granted that we do not see those that may not "reveal" their "preference" on Google but these unknowns make this paper an empirical one. A possible underrepresentation of one option is quite likely related to a following of lesser conviction and is hence not detrimental to our method.

The core difficulty consists of making sure that the searches we choose are indeed proxies for voter intent. In the Irish case this is not as easy to check by inspection as it was in Askitas

(2015a). The fact that the Irish referendum had two unrelated questions made it even more difficult. The Irish had to vote yes or no on the gay marriage question but also yes or no for reducing the age of eligibility to run for president from 35 to 21. Nonetheless what saves the day is that the gay marriage dilemma is way more engaging than the age of presidential candidate.

It is a matter of coincidence, luck and circumstance whether or not one can identify keywords that represent an opinion in Google search. In the case of a dichotomy such as a referendum it would appear as though simply taking the words "yes" or "no" ought to be enough. The closer we get to voting day the more these searches are not irrelevant random ones but related to the vote and representative of voter intent. The residual searches of random yes and no searches ought to be evenly distributed and get smaller and smaller as we approach a singular event as a referendum and ought to have a smaller and smaller footprint in the estimation error. As we will see such is the case indeed.

4. CALLING THE OUTCOME

Figure 2 depicts hourly data and 24 hour moving average data of the percentage of searches containing "vote yes" over all the searches containing either "vote yes" or "vote no". In other words the percentage of "yes" in the space of "yes or no". The horizontal line is draw at 62.07% which is the final outcome for the yes vote while the vertical line is drawn at 20:00hrs UTC on the evening voting ended. Two things jump out of the graph. The first one is that up until the late evening of May 18 the two votes as proxied by the search data were both at about 50%. The second is that late in the evening of May 21 the moving average jumped above the final outcome.

5. CONCLUSIONS

Provided an identification strategy exists and can be checked by some means, either in a more vague way as here where I look at "vote yes" or "vote no", or explicitly by braking down the searches using the exclusion mechanism as in Askitas (2015a), predicting the outcome of a referendum with Google Trends in a country like Greece (with internet penetration at 58%) or Ireland (with penetration 82%) is feasible and in fact the method I present returns eerily precise results even when traditional polling does not pick up the outcome. The difference between a classical survey and Google search (which I view as a continuous irregular panel survey of utterances) is that in the former a question is well posed but the answer may be misleading as the voter may answer strategically refusing to reveal her true preferences while in Google search the voter reveals the true answers which are however not always easy to match to the question of interest. Nonetheless Google Trends is an invaluable tool for getting information as it is happening and social science without it will soon be unthinkable.

REFERENCES

Askitas, N. (2015a): "Calling the Greek Referendum on the Nose with Google Trends," *SSRN*.

——— (2015b): "Trend-Spotting in the Housing Market," *IZA Discussion Paper Series, DP 9427*.

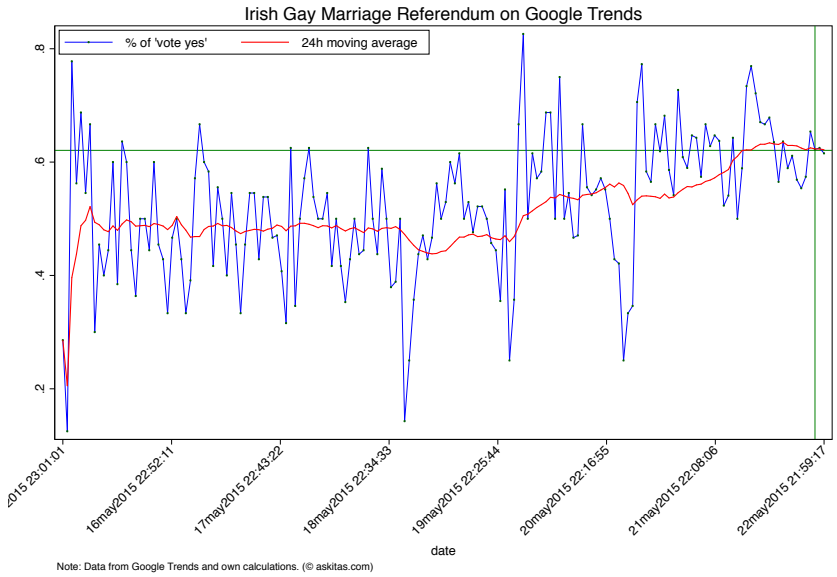

FIGURE 2.— The Irish referendum on Google Trends. Hourly percentage of "vote yes" search among searches of either "vote yes" or "vote no". The end of the series matches the final outcome with remarkable accuracy.

Data Source: Google Trends (www.google.com/trends).