

RatSWD Working Paper Series

www.ratswd.de

RatSWD ■
German Data Forum

249

Calling the Greek Referendum on the nose with Google Trends

Nikos Askitas (IZA)

July 2015

SPONSORED BY THE



Federal Ministry
of Education
and Research

Working Paper Series of the German Data Forum (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD or of the Federal Ministry of Education and Research.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD

(since 2014 Regina T. Riphahn; 2009-2014 Gert G. Wagner; 2007-2008 Heike Solga)

Calling the Greek Referendum on the nose with Google Trends¹

Nikos Askitas (IZA)²

Abstract

In a bold and risky political move the Greek prime minister Alexis Tsipras called for a referendum on June 27 2015 quitting ongoing negotiations with Greece's creditors in Brussels. The referendum framed as a yes or no question asked the Greek voters to decide whether or not they approve or reject the latest take-it-or-leave-it proposal for "program continuation" by Greece's creditors. What followed was a chaotic week leading to the referendum with intense campaigning by the two camps. Due to tense debates and increasing polarisation it became increasingly impossible to rely on traditional polling. Even the first exit polls (performed by phone on Sunday evening) could only see a marginal lead for one or the other vote at different times. Quite possibly people were jumping party lines and were unwilling to reveal their preferences. Using Google Trends I could tap into voters' true and unbiased revealed preferences and nowcast hourly what the ratio of the No vote to the Yes vote is and called an over 60% No vote well ahead of the closing of the voting urns. In this paper I document this nowcasting exercise.

Keywords: nowcasting, Greek Referendum, referendum, exit polls, Δημοψήφισμα, ναι, όχι, complexity, behaviour, data science, computational social science, complex systems.

¹ Draft of July 15, 2015. I would like to thank Konstantinos Tatsiramos for a very useful critical reading of the manuscript. Opinions and errors are mine.

² Institute for the Study of Labor, Schaumburg-Lippe-Strasse 5/9, D-53113 Bonn, Germany, nikos@iza.org;
<http://www.iza.org/home/askitas>

1. Introduction

On the 25th of January 2015 the left-wing party of SYRIZA, campaigning on an anti-austerity platform, won the general elections in Greece and quickly formed a coalition government with, what many perceived to be an unlikely partner, right wing ANEL. On February 20, 2015 a first preliminary agreement was reached between the new Greek government and the country's creditors which was based on the now infamous concept of "creative vagueness" (language which would allow all sides to claim a victory while crafting a "honorable" compromise). The months which followed proved that belief to be overly optimistic if not flat out wrong. Turning that first agreement into a concrete joint plan proved impossible as it revealed the unbridged differences between the two sides. Ever since we have been witnessing long and fruitless negotiations between the Institutions formerly known as "Troika" and the Greek government.

Negotiations came to an end when Greek prime minister Alexis Tsipras claiming an impasse in the negotiations surprised friends and foes, quit the negotiations and flew home from Brussels on June 26. In the wee hours of June 27 he called for a referendum on the last take-it-or-leave-it proposal from the creditors: a vote of yes would mean that the voter accepts that last ultimatum while a vote of no would mean the opposite.

Amid concerns of constitutionality of the referendum³ and conflicting attempts to frame a no vote both as identical with an exit from the eurozone but also as the best chance to stay in it the referendum took place with remarkable stoicism on behalf of the public. In the meantime a bank run over the weekend following Tsipras' departure from Brussels forced the Greek government to enforce capital controls and a week long banking holiday, starting Monday, at a time when thousands of pensioners were expecting their pensions, which complicated matters even more. Several European representatives as well as Greek opposition parties called for a yes vote while the prime minister himself, his party and their coalition partner called for a no vote. A number of ANEL members of parliament defected from party line and contributed to a highly complex and unpredictable situation. In the meantime the country's creditors withdrew the offer which was the basis for the referendum leading many to argue that the referendum was baseless.

In this extremely complicated situation where emotion, contagion, national pride, existential anxiety and economic despair are intertwined, behaviour and consensus formation becomes a complex system which is out of reach of traditional polling, also because revealed preferences need not persist in time or need not be faithfully represented. This is the type of situation where a small slice of Big Data available at high frequencies and in a timely fashion can show its strength.

³ Due both to its short notice and the fact that the Greek constitution does not allow a referendum on fiscal matters

Using hourly data of Google searches⁴ for yes and no respectively I am able to get to the true preferences of the voter population (while polls could get to only the revealed ones) and hence identify good proxies for intent to vote yes or no respectively. This allowed me to quite accurately forecast the No vote, which was the vote whose prevalence the entire world was interested in.

Askitas and Zimmermann (2015b) compile a good portion of social science research using Internet data but to the best of my knowledge there are no forecasting papers using hourly Google Trends data. There are of course such papers using hourly twitter data for example such as Botta et al. (2015) which also studies behavioural complexity or Cunha et al. (2014) which like we do studies voting.

The rest of the paper is structured as follows. In Section 2 I describe the data while in Section 3 I discuss a selection of relevant events and explain how the data predicts the outcome. Section 4 draws some conclusions.

2. Data

Before proceeding to describe the data some background information is necessary to better explain how I arrived at choosing my time series. Around midnight of June 26 rumours surfaced of an imminent call to a referendum by the Greek government on the ultimatum the creditors had put on the table. In the wee hours of June 27 shortly before 4 AM the Greek premier A. Tsipras made the rumours official indeed. On Sunday June 28th at 4:00 AM the parliament approved the referendum: in just six days the voter body was called to cast a yes or no vote on a complex draft of macroeconomic planning. In the meantime over the weekend a bank run was unfolding as masses of people were driving around in search of ATMs which would dispense cash. Reports had it that the ATMs were drying up but at the same time trust in the media was deteriorating rapidly⁵. We were nearing the final part of the so called "left parenthesis"⁶. In the afternoon of that same Sunday reports surfaced that a combination of capital controls and a week long banking holiday is imminent. At 21:00 hrs A. Tsipras spoke on TV insisting on the referendum, *we have only our fear to fear* he said. The stage was set. Come Monday banks were closed, capital controls were in place pensioners were lining up to get fragments of their already depleted pensions and EU president J.C. Juncker was calling the greeks to vote yes *"no matter what the question was"*.

⁴ During a discussion with Nikos Tachtsidis, a staffer at the Greek Presidency, I started thinking about whether or not I could nowcast this situation. On the 5th of July at 18:59 hrs, one minute because voting ended, I tweeted: (@askitas) "My 2 cents: 60% NO is "on the table". Selection bias (picking up the young) might cause this to be off #Greferendum" pointing to a blog post of mine at <http://www.askitas.com/2015/07/05/greferendum-now/>. As it turned out the final count of the NO vote was 61.31% and no professional pollster saw it coming. The entire world was looking at Greece between June 26 and July 5 wondering what the result of the referendum will be. It continues to do so at this writing pondering over what the no vote can buy but this is a different story which goes beyond this paper.

⁵ According to the Eurobarometer as of November 2014 80% of the Greek people don't trust Television, 60% don't trust radio and 38% mistrust the internet.

⁶ A plan commonly attributed to A. Samaras the country's previous prime minister according to which he let Tsipras come to power (something which seemed inevitable at the end of 2014) expecting his left-wing government to fall apart under the impossibility of meeting the country's financial obligations come July 2015 and hence recapture power.

As polls don't work, revealed preferences are both volatile and untrustworthy, trust in the media is waning and polarisation intensifies how can we put an ear on the ground to sense what's going on? Looking at Google Trends and at searches containing *ναί* the Greek word for yes and *όχι* the greek word for no we can only conclude that the searches are becoming more intense by the day as we are approaching the referendum and they peak around mid-night. Moreover the top searches contained both *ναί* and *όχι*. All of them were of the sort yes or no referendum revealing only that people were trying to get information on the upcoming referendum.

The time series of searches containing yes without containing no and searches that contain no without containing yes: '*όχι -ναί*' and '*ναί -όχι*' were more promising. These time series appeared to indeed capture the no and yes voters as they were searches trying to find out where and when a demonstration or a concert in favour of these votes was taking place or which prominent people were for this or that vote. In other words these searches appeared to be searches by people whose mind was made up and they were trying to locate like minded ones and events or platforms where they could express themselves (this can nicely and clearly be read off of Figure 3 where the No demonstrations occur at the peak of a surge of the no searches relative to the yes searches).

The time series I looked at was exactly these '*όχι -ναί*' and '*ναί -όχι*'. I used Google Trends (<https://www.google.com/trends/>) setting the country to Greece and the time frame to "last seven days". This gives us hourly data for the last seven days and are available by the hour. The downside is that the hourly data cannot exceed the last seven days. I took however three measurements: one early on and two later and glued them together to prolong the series.

For the forecasting exercise we do not need the prolongation. We only use hourly data between 4:00 AM on June 28 at which point the parliament votes the referendum into effect and 7:00 AM on July 5 at which point voting starts. It is however useful in order to get a better more complete understanding of our time series. The time series looks as follows.

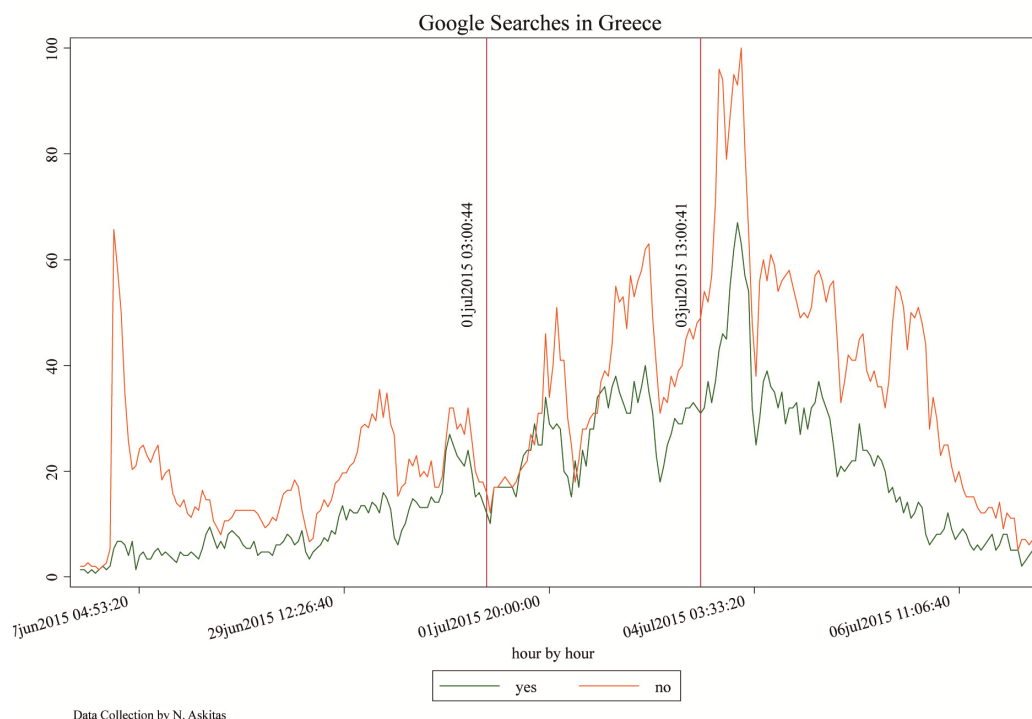


Figure 1: YES searches are “ναι -οχι” and NO searches are “οχι -ναι”

The time series is hourly and is updated by the hour. So one can really do nowcasting with hourly frequency of incoming results. It is clear from the graph that for most of the time except a few hours in the middle the NO searches were more than the YES searches. Data Source: Google Trends (www.google.com/trends).

Table 1: Top searches for NO i.e. “οχι -ναι”

βεγγος οχι δημοψηφισμα οχι οχι στο δημοψηφισμα συγκεντρωση υπερ του οχι υπερ του οχι συγκεντρωση οχι λεμε οχι δη οψηφισμα οχι οχι ψηφίζω οχι	Veggos No referendum no no in the referendum demonstration in favour of no in favour of no demonstration no we say no referendum no no I vote no
---	---

Table 2: Top searches for YES i.e. “ναι -οχι”

ναι στην Ευρωπη δημοψηφισμα ναι ναι στο δημοψηφισμα συγκεντρωση υπερ του ναι of yes συγκεντρωση για το ναι ναι στο ευρω επιτροπη υποστηριξης του ναι	yes to Europe referendum yes yes in the referendum demonstration in favour demonstration for yes yes to euro committee in support of yes
--	--

For completeness I am providing the tables of the top searches of each category in original Greek and in translation. One keyword is of particular cultural importance. The search βεργος οχι is trying to locate a scene in a greek tragicomedy involving Veggos a legendary greek comedic actor who plays a voter in a referendum shouting No while he is bombarded with Yes and physically abused.

The tables establish that it would not be a far fetched claim that our time series capture the supporters of the yes and no vote. Two issues need to be discussed however.

The first issue is that yes or no searches may also exist which are not relevant to the referendum and which are also counted in our time series. The counterargument which convinced me to continue is that firstly if there are such searches they are probably tiny compared to those relevant to such a pressing issue (as can be seen by the right hand side of Figure 1 after the vote the numbers drop close to zero) and secondly there is no reason to assume that these irrelevant searches are not distributed in some uniform way across the two options.

The second issue is that of selection. Most of the internet activity comes from around Athens (as I could see in Google Trends) and we are not capturing the poor and the old (who are most likely not online). With almost half of the Greek population living in Athens it is no surprise that it dominates internet activity. This means though that Athens is a 50% sample of Greece. Since most of the population in Athens are internal migrants with ties, friends and family to their place of origin we can easily view them as sampled from across Greece. The remaining problem that we are probably not capturing the old and the poor (who are offline) is resolved with a more cynical argument which is however rooted in reality. The old and the poor in the midst of unpaid and depleted pensions and wages, massive unemployment, capital controls and bank holidays are also more likely not to be able to afford going to vote especially if they vote away from their place of residence at their place of origin as is often the case with internal migrants in Greece.

It is well known that the Google Trends data is relative data in other words within the time unit i we are aggregating we take the number x_i of searches for a keyword and divide that by the total number of searches T_i in the same time frame, so that we form x_i/T_i . Moreover if we are observing a certain time period in our case hourly data for seven days then $i = 1 \dots n$ for some n . If then $M_n = \max_{i=1\dots n} \{x_i/T_i\}$ then the time series we get from Google is:

$$G_i = \frac{100 \cdot x_i}{T_i \cdot M_n}$$

or setting $cn = 100/M_n$

$$G_i = \frac{x_i}{T_i} c_n$$

It is obvious that two intervals of observation (i.e. two different 7 day stretches in our case) can nonetheless be grafted together by just scaling the numbers accordingly in case they have overlapping data points. Even disjoint observation intervals can be glued together by

taking differences of natural logs $\ln(G_i) = \ln(x_i/T_i) + \ln(c_n)$ because when we take their differences ($\ln(G_{i+1}) - \ln(G_i) = \ln(x_{i+1}/T_{i+1}) - \ln(x_i/T_i)$) we eliminate the scaling constants.

3. Nowcasting

Let $Y_i = c \cdot y_i/T_i$ (resp. $N_i = c \cdot n_i/T_i$) be the series of hourly search intensity for "yes" i.e. "ναι -οχι" (resp. "no" i.e. "οχι -ναι"), where y_i (resp. n_i) are the absolute numbers of yes searches (resp. no searches) and T_i are the total searches in hour i , while c is the scaling constant explained at the end of the previous section. We form the ratio

$$q_i = N_i/Y_i = n_i/y_i$$

hence we observe the ratio of absolute numbers of no searches to yes searches. Assuming these are proxies of voting intention and direction, that they all do vote indeed and that they are all realised in the voting urn in a valid vote we then have at time i that the percentage NO_i of no vote must be

$$(1) \quad NO_i = 100 \frac{q_i}{q_i + 1}$$

We used the technique of dividing two Google Trends time series in Askitas and Zimmermann (2015a) in order to find search intensity relative to other common searches rather than to the total search volume. What is unique here is that the yes and no votes are a complete partition of the total vote and hence the ratio allows us to nail the percentage of No vote. In the rest of this section I would like to discuss two graphs and to explain how we can predict the no vote on the nose. In Figure 2 I have worked out a calendar of relevant events which may explain what was driving the No vote. Notice that before Tsipras announced the referendum (far left hand side) the no searches have a huge spike. The numbers become interesting and valid for our analysis only after the parliament approves the referendum at 4 AM on June 28. Before that there may have been searches by mostly people who did not want the referendum in the first place (the sentence 'οχι στο δηλωτηφισ1α' which dominated the volumes early on can mean either of "no to the referendum" or "vote no in the referendum" depending on context and whether or not we are before the referendum is official or not).

Notice that the No vote is almost everywhere ahead except after the EFSF program for Greece had expired at midnight of June 30 and before the Eurogroup rejected a conciliatory last proposal on the part of Tsipras and his speech on public TV in which he asked people to *close their ears to the threats*. Barring a long con on the part of the creditors where they were appearing to favour a Yes vote in order to cause a No vote it can be argued that Tsipras' move to file a last compromise (even if it was a bluff) was the decisive political move and its rejection by the creditors a tactical mistake. The No vote kept gaining ground after that and it never looked back.

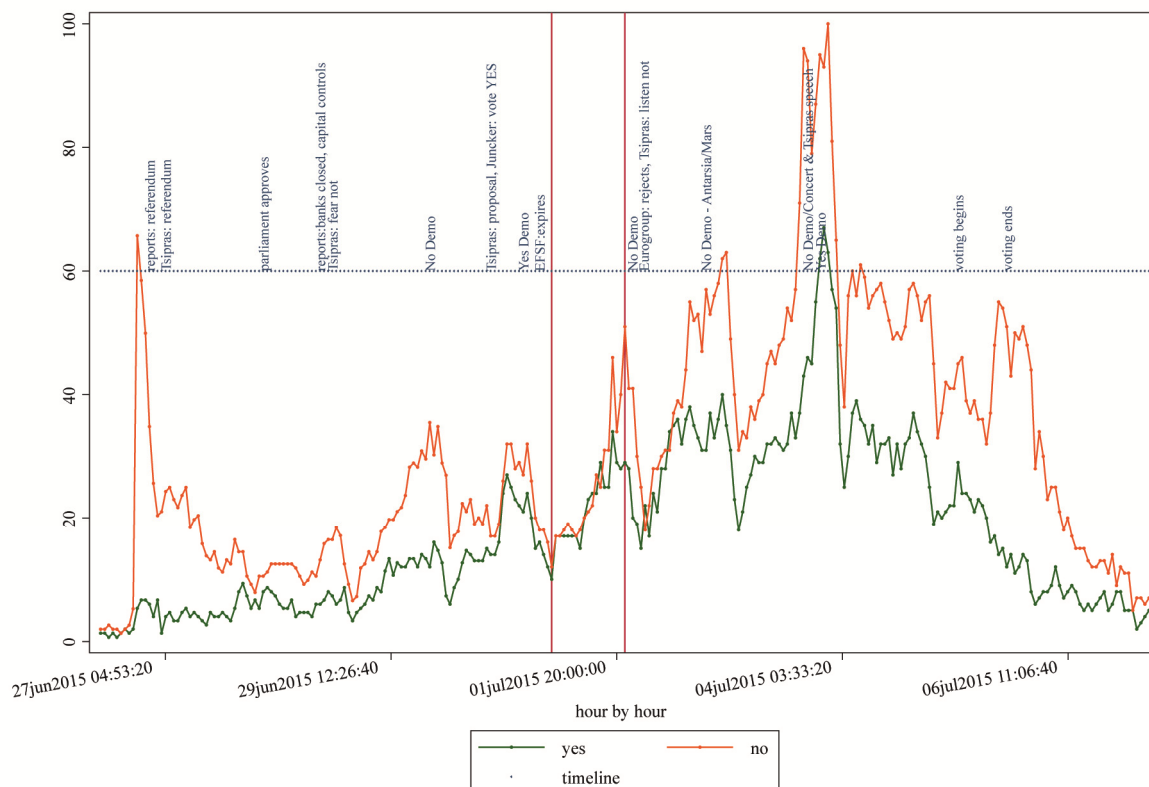


Figure 2: YES searches are “ναι -οχι” and NO searches are “οχι -ναι”

The YES vote was always trailing. The NO voters blinked only after the EFSF program for Greece expired and before the Eurogroup rejected one last compromise by A. Tsipras and Tsipras spoke on TV and said to the Greeks to close their ears to threats. Data Source: Google Trends (www.google.com/trends).

Figure 3 is the one from which we can actually nowcast the outcome. It is this graph which allowed me to call the No vote at 60% or more. It can be seen that the ratio of No to Yes vote (searches) has been dropping ever since the referendum was voted into effect. This was a terribly intense time for the population with constant flow of admonitions, with banks shut, capital controls and ATMs emitting cash at 60 Euros per debit card per day. As soon as European officials started criticising the last conciliatory proposal by A. Tsipras the ratio is surging again and after his proposal was rejected and he appeared on TV he had locked the victory and he only had to play the game to the end without any mistakes. Hours after the moving average of the No to Yes vote surged above the 1.5 line it never dropped below again. The value of 1.5 is simply the maximum ratio that was upheld for the longest time before voting started which however local minima of the hourly time series kept approaching from above again and again without ever reaching or crossing (right hand side of Figure 3).

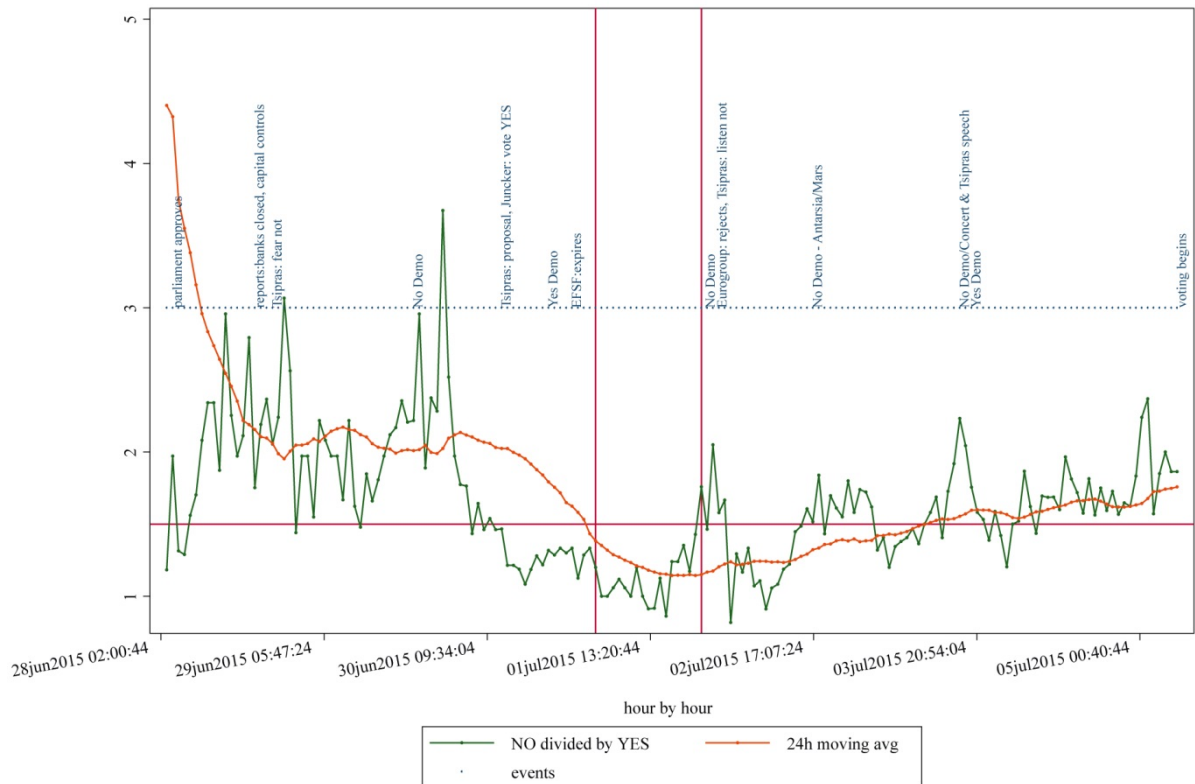


Figure 3: YES searches are “ $\nu\alpha\iota$ - $\omicron\chi\iota$ ” and NO searches are “ $\omicron\chi\iota$ - $\nu\alpha\iota$ ”.
The ratio of No to Yes votes and its 24 hour moving average.

The moving average bottoms out on July 1st at 1.143911 and increases ever since crossing the line $y = 1.5$ on July 3 at 13:00 hrs. It never dropped below 1.5 after that point. Data Source: Google Trends (www.google.com/trends).

The reader can now figure out that setting $q_i = 1.5$ in equation 1 produces $NO_i = 60^7$. So to summarise. Our identification strategy was to look at searches containing yes without no and those that contain no with yes. We established that the top searches are a) relevant to the referendum and b) partial to the corresponding vote. Moreover by looking at such searches before the referendum emerged and days after it was completed we established that those of the yes and no searches that had no relation to the referendum were of insignificant volume (besides the reasonable assumption that they must be uniformly shared). Hence our two time series are reasonably good candidates for proxies of the corresponding vote and our empirical exercise proved this to be right.

⁷ I tweeted my prediction to @mark carrigan at 10:26 on July 5 and shortly after that I emailed K. F. Zimmermann my prediction. I tweeted it publicly at 3:39 PM, to @D Blanchflower at 5:41 PM and at 18:59 PM one minute before closing of voting I tweeted that am calling it “60%” for No. At 7:23 PM Greek TV reported that based on 7.4% of the vote the No vote comes slightly under 60%. Nikos Tachtsidis whom I also sent the result to one minute before closing of the vote and who, in disbelief, replied that exit polls are calling it only marginally for No contacted me soon after the first TV report placed the no vote close to 60% to congratulate me at which point I doubled down: *it is going to settle well above 60%* I said. The final result came in at 61.31%. I have also produced 4 relevant blog posts on my site at <http://www.askitas.com>

4. Conclusions

This paper is the first to our knowledge to use the hourly "last seven days" option of Google Trends and I suggest that this option ought to be explored more by researchers. I used a small slice of big data and a lot of intuition (in lack of better access to search microdata) to create a simple live nowcasting of the Greek referendum. The simplicity of the yes/no vote and the fact that people wanted to join demonstrations in support of their own conviction worked to my advantage and hence I could nicely capture the ratio of one vote to the other. This paper is a small data vignette to demonstrate that in a situation of breaking news and wild dynamics where behaviour is complex and professional polling is unable to perform well I could forecast the result of the referendum which took place in short notice and in highly volatile circumstances on the nose.

References

- ASKITAS, N. and K. F. ZIMMERMANN (2015a): "Health and well-being in the great recession", *International Journal of Manpower*, 36, 26–47.
- (2015b): "The internet as a data source for advancement in social sciences", *International Journal of Manpower*, 36, 2–12.
- BOTTA, F., H. S. MOAT, and T. PREIS (2015): "Quantifying crowd size with mobile phone and Twitter data", *Royal Society Open Science*, 2.
- CUNHA, E., G. MAGNO, M. A. GONALVES, C. CAMBRAIA, and V. ALMEIDA (2014): "He Votes or She Votes? Female and Male Discursive Strategies in Twitter Political Hashtags," *PLoS ONE*, 9, e87041.