1. introduction

The UK, as most other world economies, is facing a period of Covid-19-driven hardship. In such a reality, immigration is likely to become a sensitive topic in policy making. Particularly so, given that the UK’s exit of the European Union was in part driven by the desire to pass immigration-related legislation on UK’s terms.

Existent debates are manifold. While the desire to define own policies partially stems from a fear of an influx of immigrants (Source/some Brexit campaign), other argue that preventing immigration of skilled workers is likely to hinder the country’s economic recovery due to gaps in currently-high-demand-jobs (Grierson, 2020). In any regard, depending on how circumstances, also beyond COVID-19, will develop in other countries, people may seek to leave their homes in the search of a better future. Hence, debates on immigration are and will likely remain of core parliamentary importance.

As discussions in the House of Commons (HoC) are instrumental to the unfolding of policies and pieces of legislation regarding immigrants, for the individuals’ better or worse, understanding politicians contributions to the debates can provide valuable insights into how certain policies came about. This approach uses speech data to investigate sentiments and themes about immigration with the intention of unfolding how some issues are framed (Bara, Weale & Bicuelet, 2007).

This blog aims to offer a historical overview on how immigration was and is discussed in parliamentary debates, in addition to the main themes and party sentiments regarding specific topics of migration-related discussions. We believe such analysis can provide politicians and policy makers with useful insight that can help them generate an actionable strategy for coalition building for immigration related policies.

- should probably mention the two main events in Introduction and why we focus on them

2. Data

To understand how debates related to immigration are framed and perceived when discussed in British parliament, we used the HoC-related dataset from a database called ParlSpeech V2 by Christian Rauh and Jan Schwalbach (2020). This dataset (as the entire database itself) is unique in its scope, covering all parliamentary debates from 1998 and up until 2020, resulting in 1,956,223 speeches (Rauh & Schwalbach, 2020, p. 10). In this regard, speeches represent individual contributions by members of parliament, which were collected from the digital Commons Hansard that contains the plenary protocols and documents from which speech texts and metadata are extracted. As a result, the corpus contains a range of covariates like *party* *affiliation* and *agenda,* which facilitate a detailed set of context and party-specific analyses. For that end, we also leverage the (established to produce reliable estimates) Lexicoder 2015 sentiment dictionary that consists of 2,858-word patterns relating to negative sentiment and 1,709-word patterns, indicating positive sentiment (Young & Soroka, 2012). This sentiment dictionary is particularly relevant to our purposes as it was designed to analyse sentiment in political language of legislative speech [[1]](#footnote-1) and has been applied specifically to migration discourse. [[2]](#footnote-2)

2.1 Subset

Choosing a subset for analysis is a challenging task, and in our case, the decisions we took were related both to substantive and practical consideration of needing to narrow down a very large database to perform a more in-depth analysis. In a first step, we choose to focus on texts from 2010 to present day, leaving us with slightly less than 750.000 individual contributions. 2010 is a good starting point for our analysis because that was the year of the Tory manifesto and the general elections which resulted with a win for the Conservative party. This allows a sufficient time frame that has observations both before our main events of interest, namely the 2015 General Election, the migration wave and the Brexit Referendum, and after, from 2016 until 2020. In terms of content, we subset the corpus to those contributions that either contain a reference to keywords related to our topic of analysis, or were made as a response to agenda points that contain such keywords. Specifically, the keywords “I/immigra\*”, “R/refuge\*” and “A/asylum” were used. As a continuation of existent research, we expect parliamentary debates to be explicit in their language, meaning that if immigration is discussed, one of these keywords will show either in the agenda description or in the speech itself, and would therefore allow us to capture most of the substantive debates regarding immigration (Van Dijk, 2000). In a final step and within the content-based subset, we excluded contributions that were shorter than 10 words, as well as included only the 5 parties with largest overall contributions made. Namely, Conservatives, Labour, Liberal Democrats, SNP and DUP. This type of subsetting allows us to focus our analysis and remove noise from unrelated text, and yet, contain the limitation of not including any documents who discuss immigration without mentioning the three key terms chosen in either agenda description or text. Further, by this subsetting we are very likely to lose short responses to speeches carried out.

The aforementioned steps yielded a final subset of 22.257 individual contributions, representing about 3% of the parliaments overall debates during that period as well as about 6.25% of the overall time spend in debates. Further, the parties selected represent about 98% of the overall contributions made.

2.2 Foundational Dataframes & Considerations

There are two main threads to our analysis. Firstly, we use a general subset of the HoC parliamentary debates, described above. Secondly, we use a more targeted subset that only capture the immediate context surrounding the selected keywords. Essentially, this subset contains bubbles of words found in range of 20 before and after a keyword. Importantly, if two keywords are presented within such a bubble, there would be no multiplication. We created this subset in order to be able to pursue a deeper analysis of how these terms are used. By looking at these words in their context, we can grasp better the various meanings attributed to them and see what happens in these moments in which key terms are explicitly mentioned.

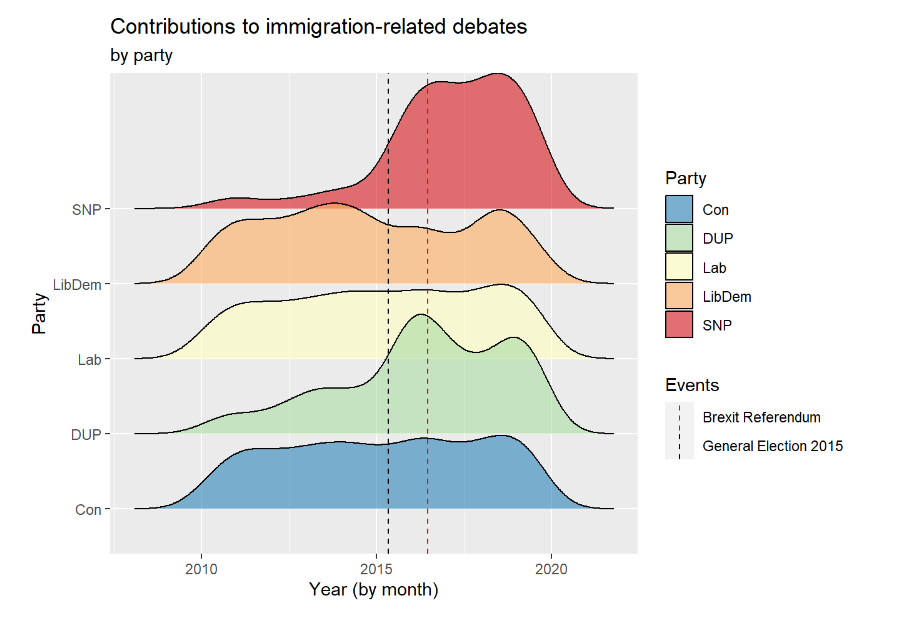
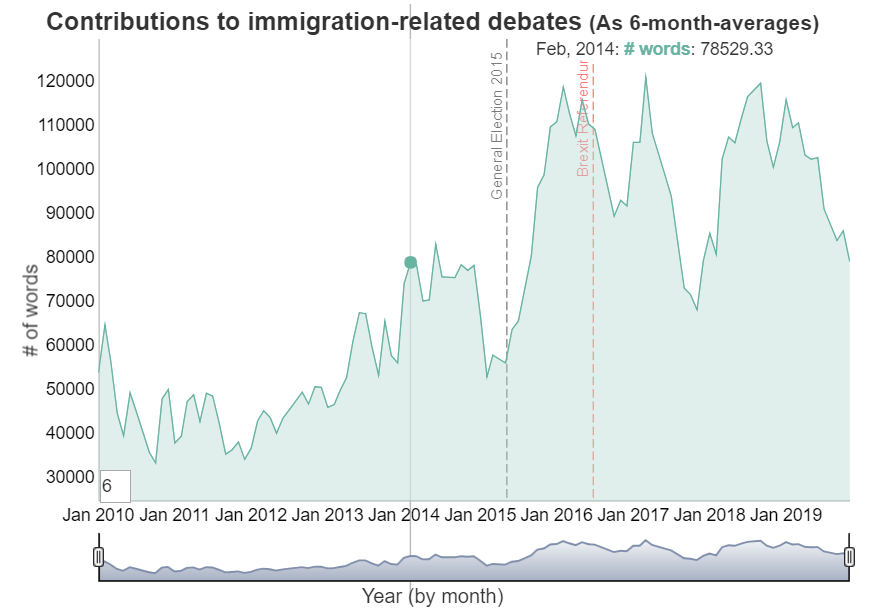
mention events etc.

In order to constrain the analysis further, we take the 2015 general election and Brexit as two fulcrums for the analysis. This approach assumes that understanding what happened in past political and social disruptions can assist policy makers in devising strategy to deal with future, or already present turbulent periods

.

3. Descriptives | General Findings

Graphs that would go in this section:

The first density plot depicts frequency (y) of individual contributions (regardless of their length) over time. Technically speaking, this equals the total count of documents for each month between 2010 and 2020. From this, one can infer the overall prevalence of immigration-related debates in the HoC between 2010 and 2020, irrespective of party. Evidently, following the general elections in 2015, there was a rapid increase in discussion about immigration, which is aligned with the progression of the vast immigration wave across the EU. The spikes and breaks are likely due to the different recess dates the HoC has.[[3]](#footnote-3)

*Plot 2* depicts the amount of unique agenda points either dedicated towards immigration related matters. What becomes clear is that the overall amount of agenda points devoted or somehow related to immigration has almost tripled between 2010 and 2020, with a nearly linear increase over the years.

*Plot 3:* Prevalence of immigration debates over time by month | Total number of words as a proxy for time spent on debating.

Different from the previous plot that ignored length of speakers’ contributions, this plot uses the sum of words used within debates as an indicator of the time spent on the respective debate. Considering the HoC only has a limited time available to discuss agenda points, devoting more time towards a debate may indicate certain priorities. In this regard, plot 3 depicts the 6-month-average total amount of words spend on immigration-related debates. By looking at the 6-month averages, it is possible to observe whether debate-preferences prevailed over time or whether they only peaked for short periods.[[4]](#footnote-4)

From January 2012 to November 2014 we are able to observe a steady increase in time spend on debates with regards to their 6-month averages. This is likely due to the spikes showing on a monthly level in both January and June of 2014. The second half of 2014 as well as the first half of 2015 saw less time being devoted to immigration related debates. This suggest that overall, the content on which we selected as our subset did not increase in particular prevalence before the 2015 General Election.

However, between May 2015 and June 2016, hence the year following the general election and leading up to the Brexit referendum, there was a major increase in time spend on immigration-related debates. On average, the HoC spend almost twice as much time on immigration related debates during Sep 2015 - February 2016 when compared to the period of December 2014 - May 2015. Hence, debates seem to have gained in priority after the General election and leading up to the referendum.

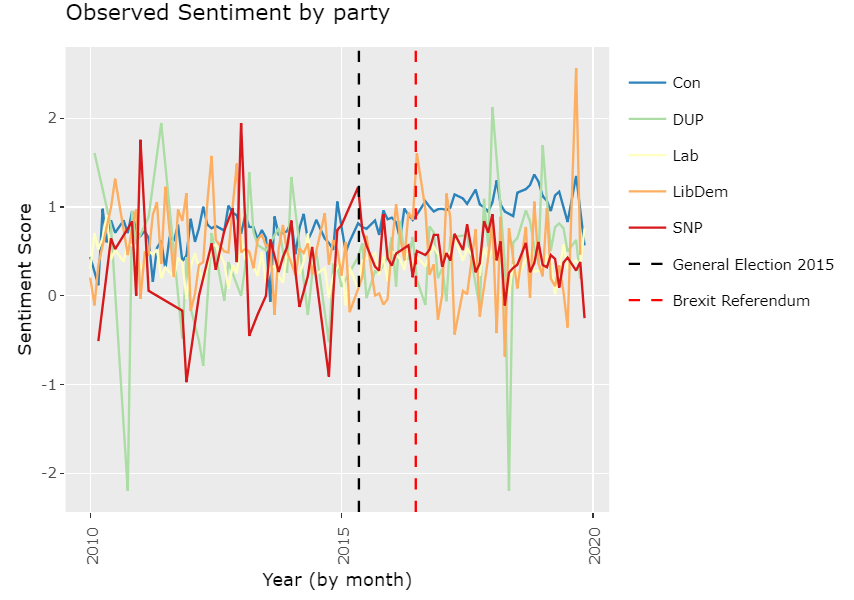
Plot 4. Number of words spent on immigration related debates (by party)

Zooming in further, this density plot gives us a sense of the frequency each party discussed immigration related issues every month during the time frame investigated. What it shows is that while the SNP and the DUP spoke more about immigration after Brexit, other parties has a more constant trend of engagement with immigration related speech. Importantly, the information that can be gathered from this graph is limited in that it does not tell us anything about substance of these speeches, but crudely how many words were used. Nevertheless, this descriptive visualization does help us get an initial sense about the prevalence of immigration related speech in each of the parties we are focusing on.

## 4. Sentiment

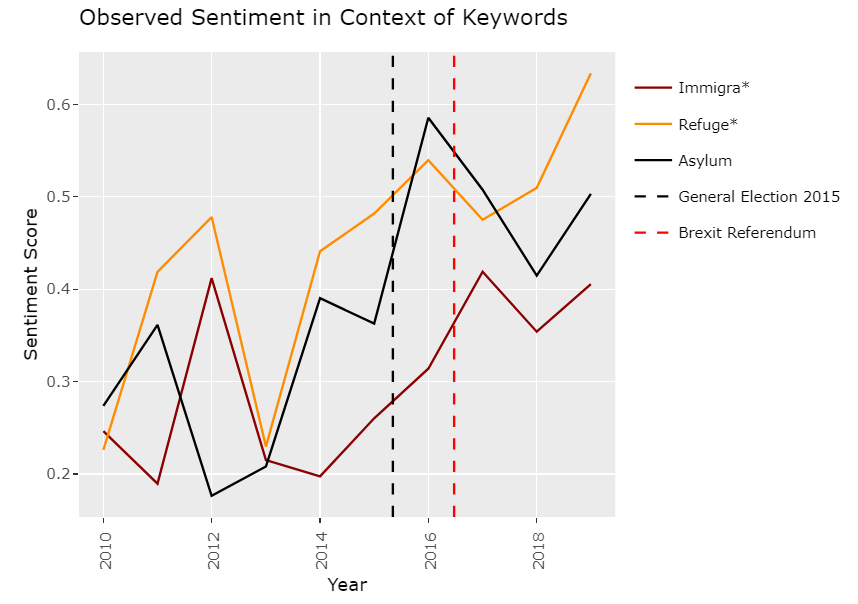
## Short description of the method and package used:

Graph: Sentiment by party



\*I still think that we should take out these two graphs because they capture sentiments from debates which are very different and thus does not tell us much or what it does tell us is likely to have significant "measurement error".

## 4.1. Sentiment in Keyword-specific context



Thoughts about sentiments:

We have a problem that we are not controlling for documents’ length meaning that in periods where there are relatively more contributions we would also see “increase” in sentiments. 🡪 would be good to find a way to normalise debate size and see what happens.

## 6. Topical Analysis

In order to better understand developments in migration-related debates over time, the different party positions in these debates, and the sentiment associated with discourses of migration, it is important to distinguish these debates by the themes they discuss. To do this, we use the Structural Topic Model (STM).

- topics

Using the STM package in R, we model 6 topics from the content of migration-related documents and assign each document a theta score for each topic. These scores represent the proportions of prevalence of each topic for each document. Next, we test how exclusive and coherent these topics are. While 6 topics may seem few, we are modelling topics from a subset of parliamentary debates that use migration-related keywords. This already limits the extent of topics potentially covered by these debates. Our test also shows that 6 is a sufficient number for allocating exclusive and coherent topics.

-topics dataframe

We combine the topic scores of each document to our dataset of migration-related debates. Next, we attribute to each topic a name based on the first 3 FREX terms, the words that are more frequent and most exclusive to each topic. One topic, labelled “allowance, tax, dwp,” may refer to content on migration that related to social welfare. The topic named “eea, seasonal, visa” relates to the portion of debates that concern the economic dimensions of migration. “Iraq, Rohingya, Libya” describes the portion of debates that mainly concern refugees fleeing conflict. The topic “tb, Sikh, Auschwitz” seems to indicate a collection of less prevalent migration-related topics such as Tuberculosis, specific migrant communities, and the Holocaust. The title “unaccompanied, trafficked, detention” describes the most vulnerable populations of migrants and the humanitarian concerns of migration. Finally, the topic “vote, voting, motion” includes the procedural vocabulary of the House of Parliament.

-topic\_by\_party\_plot

We plot the proportions of topics covered by each party over all the years of debate (2010-2020). This plot illustrates two main things. First, it demonstrates the relative prevalence of these six migration-related topics in parliamentary debates. From this we can see that aside from procedural vocabulary, the three most prevalent topics in these migration debates are economic migration, refugees fleeing conflict, and the humanitarian concerns of migration.

In addition, this plot compares how different parties discuss migration, in terms of these six topics. Here we see that the Conservatives discuss economic migration more than any other major party and to a greater extent than they discuss refugees fleeing conflict or the humanitarian concerns of migration. In comparison to the Conservatives, the Labour party discusses migration more often in the context of social welfare and humanitarian concerns.

-topic\_by\_party\_and year

What is noticeable from this graph is that parties converge in how much they discuss topics towards the 2015-2016. Three explanations may shed light why we see this trend. The first reason relates to external developments with some influence on the UK in this time period. Noticeably, the progression of the refugee crisis. Hence, new, more pressing large-scale *national* issues enter the agenda, irrespective of party. Secondly, as usually occurs towards general election, all parties discuss broadly similar agenda points which are found in the core of the political discourse. Lastly, Brexit which completely changed the context on a national level in which policy is discussed. Some topics show a party convergence even after Brexit, such as economic migration, because it then become a really important topic for the whole UK and all parties now. One exception to this may be the Scottish party, which makes sense, as they still want Scotland to be part of the EU and therefore have a different sentiment towards economic migration.

-topics\_by\_time\_plot

To explore the change in topic prevalence over time, we plot the yearly average prevalence of each topic between 2010 and 2020. For reference, the vertical, black, dashed line marks the 2015 election and the beginning of the so-called Syrian refugee crisis in Europe, while the red, dashed line represents the BREXIT referendum in 2016. This plot shows that while the topic of social welfare enjoys relatively little prevalence in migration-related debates, it was most prevalent in the earlier years of the decade and has begun regaining attention in recent years. We also find that the topic of economic migration experienced a sharp decline in attention after 2013 has only risen to prominence once again in post-BREXIT debates. The topic of refugees fleeing conflict first received attention in 2011 (in the context of civil war in Libya) before becoming nearly three-times more prevalent at the peak of the Syrian refugee crisis in 2015. The topic lost traction in 2016 after the BREXIT referendum shifted political attention to other concerns regarding migration. (why did humanitarian concerns peek in 2016? Also related to events connected to the Syrian refugee crisis and passage to Europe?)

-topics-sentiment correlation

* Describe correlations of topic and sentiment: what are we doing here
* Describe findings…

Grierson, J. (2020)

<https://www.theguardian.com/politics/2020/dec/15/post-brexit-key-worker-shortage-may-hamper-uk-economic-recovery>

Heidenreich, T., Eberl, J. M., Lind, F., & Boomgaarden, H. (2020). Political migration discourses on social media: a comparative perspective on visibility and sentiment across political Facebook accounts in Europe. *Journal of Ethnic and Migration Studies*, *46*(7), 1261-1280.

Bara, J., Weale, A., & Bicquelet, A. (2007). Analysing parliamentary debate with computer assistance. *Swiss Political Science Review*, *13*(4), 577-605.

1. <http://www.snsoroka.com/data-lexicoder/> [↑](#footnote-ref-1)
2. (Heidenreich et al., 2020) <https://www.tandfonline.com/doi/full/10.1080/1369183X.2019.1665990> [↑](#footnote-ref-2)
3. For example, in the months before January 2018, where there seem to be a decrease in debates, there were several holidays, including summer holiday, an external conference, November break and Christmas holiday. https://www.parliament.uk/about/faqs/house-of-commons-faqs/business-faq-page/recess-dates/list-of-previous-commons-recess-dates/ [↑](#footnote-ref-3)
4. To give you an example, looking at the number of words on Dec 2011 indicates the monthly-average amount of words spend on immigration related debates during the second half of 2011. [↑](#footnote-ref-4)