The UK, as most other world economies, is facing a period of covid-driven hardship. In such a reality, immigration is likely to become a sensitive topic in policy making. Firstly, some argue that the preventing immigration of skilled workers is likely to hinder the country’s economic recovery due to gaps in currently-high-demand-jobs. And secondly, depending on how things would develop in other country, people may seek to leave their homes in the search of a better future.

As a political Think-tank, we were approached by the ---?--- to provide them with an historical overview on how immigration was and is discussed in parliamentary debates, what are the main themes and what are parties’ sentiments in regards to these discussions’ topics. They believe such analysis can provide them with useful insight that can help them generate an actionable strategy for coalition building for immigration related policies.

As the discussions in the House are instrumental to the unfolding policies and pieces of legislation regarding immigrants, understanding what politicians say in the debates can provide valuable insight into how some policies came about. This approach uses speech data to investigate sentiments and themes about immigration.

2. Packages and data

explain which data has been used and forms the basis of our analysis – text analysis, what it is, text sentiment and topic and what can we get out of it.

To understand how immigrants are framed and perceived when discussed in parliamentary debates, we used a database called ParlSpeech V2 by Christian Rauh and Jan Schwalbach (2020). This database 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). The text was collected from the digital Commons Hansard that contains the plenary protocols and documents from which speech texts and metadata are extracted. The corpus contains a range of covariates like party affiliation and agenda which facilitate better analysis of the various ways in which the topic of consideration is discussed by the different parties’ representatives and depending on the agenda context. 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).

2.1 Subset

explain the subset + limitations Choosing a unit 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. Thus, we choose to focus on texts from 2010 to present day. 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 allow us for 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 only to those speeches that contain a reference to key words related to the topic. Specifically, “immigra”, “refugee” or “asylum” because we expect parliamentary debates to be explicit in their language, meaning that if immigration is discussed one of these key words will show either in the agenda description or in the speech itself and therefore we think this method would allow us to capture most of the substantive debates regarding immigration (Van Dijk, 2000). 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.

2.2 Foundational Dateframes & Considerations

mention that we will look at two basic subsets: One general one with all obrservations of the initial subset, and one based on the context of the keywords. Justify why.

mention events etc.

*2.2.1 General Corpus*

*2.2.3 General consideration/definitions used across analysis*

3. Descriptives

Justifications and thoughts here.

* Subset covers 3.05% of total debates in that time period and 6.35% of the total time spend in debates.
* This section looks at the overall prevalence of immigration-realted debates in the HoC between 2010 and 2020, irrespective of party.
* We use a density plot that depicts frequency (y) across time.
* Any reference is based on the subset and therefore immigration-related.
* Plot 1 shows the number of individual contributions made over time. Technically speaking, this equals the total count of documents for each month between 2010 and 2020. Substantively, one document reflect one individual’s contribution irrespective of its length, tone etc.
* Findings plot 1: Need to mention spikes and breaks (likely due to the different phases of the HoC; e.g. summer breaks)
* Plot 2 depicts the amount of unique agenda points either dedicated towards or somehow relating to immigration. Substantively, this means that each agenda point, irrespective of its lengths, will be counted.
* Findings plot 2: 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. While plot 1 shows the overall count of unique contributions to immigration-related debates, it does not give substantive insights into the lengths of those contributions. We argue that looking at the overall amount of words used within debates is a relatively clear indicator of the time spent on the respective debate. This is important, as the HoC only has a limited time available, 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, we are able to observe whether debate-preferances prevailed over time or whether they only peaked over a short time. 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.
* Findings plot 3: While sharp ups and downs were still visible in plots 1 and 2, averaging over 6 months allows for a smoother observation of debate evolution. 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 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, saw 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 - Feb 2016 when compared to the period of Dec 2014 - May 2015. Hence, debates seem to have gained in priority after the GEneral election and leading up to the referendum.

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

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

This density plot gives us a sense of the frequency each party discussed each party discussed each month during the time frame of our research. Basically, what it does it counts how many words each party each party invested in speaking about immigration related topics. So, for example, while the SNP and the DUP spoke more about immigration after Brexit, other parties exhibit 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

Sentiment | Overall Corpus

Graph 1: Overall Sentiment

Graph 2: Sentiment by party

\*I still think that we should take out these two graphs because they captures 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".

## 5. Sentiment in Context

2.2.2 Create KWIC - Dataframe, Corpus and Dfm

Subset KWIC according to keywords

Sentiment Keywords in Context of keyword

Graph 3: KWIC sentiment

Thoughts about sentiments:

1. 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.

We can be realists or anti-realists about topics → Anti-realism: topics are ‘lenses’ → Realism: topics are real discourse units, e.g. themes, categories, etc.