

**Attitude towards Immigrants: Evidence from U.S. Congressional  
Speeches**

Neha Bose

[\(This paper also appears as CAGE Discussion Paper No: 464\)](#)

March 2020

No: 1259

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

# Attitude towards Immigrants: Evidence from U.S. Congressional Speeches\*

Neha Bose<sup>†</sup>

March 26, 2020

## Abstract

Immigration and attitudes towards immigration have been key features in economic development and political debate for decades. It can be hard to disentangle true beliefs about immigrants even where we have seemingly strong evidence such as the voting records of politicians. This paper builds an “immigration corpus” consisting of 24,351 U.S. congressional speeches relevant to immigration issues between 1990-2015. The corpus is used to form two distinct measures of attitude towards immigrants - one based on sentiment (or valence) and one based on the concreteness of language. The lexical measures, particularly sentiment, show systematic variation over time and across states in a manner consistent with the history and experiences of immigrants in the USA. The paper also computes a speaker specific measure of sentiment towards immigrants which is found to be a significant positive predictor of voting behaviour with respect to immigration related bills. Applying a Latent Dirichlet Allocation (LDA) topic modelling algorithm provides further insight into how different topics (such as border security or national security) have risen and fallen in importance over time in the face of key events such as 9/11.

## 1 Introduction

The number of international migrants worldwide has increased rapidly in recent years with the total number of migrants reaching 272 million in 2019 ([United Nations \(2019\)](#)). Political turmoil, conflict, poverty and limited educational or employment opportunities force millions of people to migrate from their countries. The largest number of international migrants were reported to be living in the U.S. in 2019, the total number reaching

---

\*I am grateful to Daniel Sgroi and Thomas Hills for their invaluable feedback and guidance. I thank Giovanni Facchini for his generosity with time and for sharing the voting data used in the paper. Lastly, I am thankful to Clement Imbert, Sharun Mukand and seminar participants at the University of Warwick for their helpful comments.

<sup>†</sup>Department of Economics, University of Warwick, email: [n.bose@warwick.ac.uk](mailto:n.bose@warwick.ac.uk)

51 million (United Nations (2019)). Immigrants contribute significantly to the economy of the host country in terms of filling key labour market gaps, contributing to tax revenues and social security (United Nations (2017)). Particularly in the US, immigrants comprise almost a third of the workforce in several industries including farming, fishing and forestry, textile manufacturing, food manufacturing, and health care; immigrants help the ageing U.S. population, which is crucial for improving the ratio of workers to retirees as 78% of the foreign-born population are of working age compared to 59% of native born; immigrants contribute significantly to the U.S. housing market and immigrants' children tend to gain more education, have higher salaries and work in higher-paying professions compared to their parents (Sherman et al. (2019)). Yet, immigrants, as one of the most vulnerable sections of the society, are subject to constant human rights violation and discrimination.

In the US, for example, the members of the Congress play a massive role in terms of shaping legislation related to immigrant issues. Congress members vote on bills which either liberalise or restrict immigration. When predicting voting behaviour of congress legislators on immigration bills, variables such as party affiliation, regional ideologies and differences (Fetzer (2006); Goldin (1994)), the district's economic, demographic and social variables (Milner and Tingley (2011); Facchini and Friedrich (2011); Gonzalez and Kamdar (2000)) have been observed to be helpful. This paper looks at another measure which helps shed light on variation in immigrant outcomes and voting behaviour of legislators on immigration related bills - congressional speeches. The congressional speeches record all speeches verbatim delivered by legislators on the floor of the Congress.

This paper uses the United States Congressional Records from 1990 to 2015 which is a record of all speeches on the floor of Congress. From the Congressional Records, an "immigration corpus" is formed by identifying speeches which contain *immigration*, *refugees* and related words. This yields a corpus consisting of 24,351 congressional speeches with 43.6 million total number of words. The corpus comprises of speeches from 1098 unique speakers.

This immigration corpus is used to form two distinct linguistics measures of immigrant attitude - immigrant *sentiment* or valence and language *concreteness*. Valence of a word refers to the pleasant emotion conveyed by a word, with the rating increasing as it moves from unhappy to happy. Valence rating of immigration speeches acts as a measure of pro-immigrant sentiment. Concreteness<sup>1</sup> value of immigration speeches are computed as a proxy for social distance. Self-categorization theory posits that humans are conditioned to view different social groups as either in-group (within close social proximity) or out-group (socially distant) (Turner (1985)). Concrete linguistic patterns reflect positive out-group behaviour or lesser perceived social distance as supported by Linguistic Intergroup Bias

---

<sup>1</sup>Concreteness refers to a word's ability to make specific and definite reference to particular objects (Hills et al. (2016)).

model (Maass et al. (1989)) and Construal Level Theory (Trope and Liberman (2010)). Use of concrete language declines when one shifts from describing family and friends to neighbours to colleagues to foreigners (Snefjella and Kuperman (2015)). This acts as an indicator of intergroup bias which helps create and sustain stereotypes about social groups (Mastro et al. (2014)). The approach in this paper is similar to Li and Hills (2019) which forms an immigration corpus from a corpus of New York Times articles to quantify and compare sentiment and social distance towards different immigrant groups, as expressed in the articles.

The sentiment and concreteness measure of immigration speeches are used to examine variation of immigrant attitude of the U.S. congress members over time and by political affiliation and region. The variation of lexical measures are consistent with the history of immigrant outcomes, with a marked decline following the September 2001 terrorist attacks in the U.S. and the subsequent introduction of several anti-immigration bills.. The lexical measures, particularly immigrant sentiment, display strict polarization with regards to immigration, consistently since early 1990s. There is also a clear regional distinction with southern U.S. states displaying more negative immigrant sentiment and the north-eastern states displaying pro-immigrant sentiment.

The paper also uses voting data of congressional members on seminal immigration bills in the decades before and after the 9/11 attacks. The paper finds that speaker speech sentiment is a significant positive predictor of speaker's voting pattern even after controlling for a list of speaker characteristics, party affiliation and district level variables. Depending on the econometric specification, a 1 standard deviation increase in speaker speech sentiment can increase the probability of a pro-immigration vote by 2 to 5 percentage points. Speaker speech sentiment is an especially strong predictor of voting behaviour in the southern, south-western and western U.S. states.

Lastly, the paper uses Latent Dirichlet Allocation or LDA topic modelling algorithm (Blei et al. (2003)) to identify the different immigrant related topics discussed in the U.S. congress. This helps recognize the key topics for immigrant legislation and examine the distribution of the different topics over time. The distribution of topics shows a hike in border security and national security legislation following the 9/11 attacks and a decline of legislation related to educational policies and government aid in speeches addressing immigration. The distribution of topics over time expresses the variation in concern over different immigrant issues and is generally reflected by the nature of the bills introduced in the congress during the time. Applying LDA is an alternative,time saving and efficient way of examining the historical trends of immigration legislation compared to extensive manual studying of congressional bills.

Broadly, this paper contributes to the following branches of literature. Firstly, it contributes to the small branch of literature which looks at lexical bias towards immigrants in news coverage (Li and Hills (2019); Mastro et al. (2014)). Related to this is the paper

by [Rice et al. \(2019\)](#) which looks at implicit racial bias in written judicial opinions in U.S. state and federal courts. This paper also contributes to previous papers which have analysed U.S. congressional speeches ([Gentzkow et al. \(2019\)](#); [Jensen et al. \(2012\)](#)). Lastly, the paper contributes to the existing branch of literature which examines roll call voting of congress members on immigration bills ([Facchini and Friedrich \(2011\)](#); [Milner and Tingley \(2011\)](#); [Fetzer \(2006\)](#); [Gonzalez and Kamdar \(2000\)](#)).

Congressional speeches delivered on the immigration issues are crucial for the following reasons: (1) They are better able to capture the *intensity* of the speaker’s attitude towards immigration since it yields a continuous measure. This is superior to examining a speaker’s attitude towards immigration based on past voting behaviour which gives a binary measure since a vote is either in favour of or against liberalising immigration. (2) The immigration corpus, due to its sheer volume, is expected to be more informative of the speaker’s attitude towards immigrants compared to other measures, such as past voting behaviour. (3) In the absence of past voting behaviour, in case the bill does not go for a vote for example, it is a useful measure for estimating the speaker’s attitude towards the particular issue under debate.

Moreover, the study of political speeches, not only on the floor of the Congress, but also during campaign trails and other public appearances, is crucial to form a measure of the speaker’s attitude towards immigration or other relevant issues. The benefit of the method proposed in this paper is that it can be applied to study politicians’ attitude on other crucial issues such as gender topics, climate change, taxation etc. Forming a lexical measure of politicians’ attitude towards any key issue can be hugely helpful in predicting legislative outcomes related to the topic.

The rest of the paper is structured as follows: sections 2 and 3 describe the data and the methodology used in the paper, respectively. Section 4 presents the results obtained. Section 5 concludes.

## 2 Data

### 2.1 Congressional Speeches and Speaker Data

The paper uses United States Congressional Records from 1990 to 2015. This is a record of all speeches on the floor of Congress (for both chambers of the Congress). The records were digitized using optical character recognition or OCR on scanned print volumes ([Gentzkow et al. \(2019\)](#)). The source of the data from 1990 to 1993 is HeinOnline. The source of the data from 1993 to 2015 is the United States Government Publishing Office. The records were issued in bound editions compiling data from an entire congressional session. This yields a total of 1,767,180 speeches. The speakers and the state they belong to are identified at the beginnings of speeches. The speaker is matched with their biographical data

obtained from [Github repository](#) on congress legislators. The repository data contains information on current and historical U.S. legislators such as party affiliation, year of birth, gender, race, military service and ICPSR code<sup>2</sup>.

## 2.2 Voting Data

For voting records on immigration bills the paper uses the data from [Facchini and Friedrich \(2011\)](#), which gives a list of 1265 immigration votes cast by 445 congress members in the U.S. House of Representatives (Appendix A). The authors identify 8 bills from 1990 to 2006, 3 before and 5 after the 9/11 attacks, which affect outcomes for immigrants in the US. The list only includes final votes and excludes intermediate votes (such as floor amendments). The list of these 8 bills and whether the bills were for (*pro*) or against (*contra*) liberalizing immigration are presented in Table 1. Votes on immigration related issues increased substantially after the 9/11 attacks in 2001, although the effect was not immediate. As the list shows, votes on immigrant rights and benefits peaked around 1995-1996, only to be followed by a stark increase in votes on border security after 9/11 in 2005-2006 ([Fennelly et al. \(2015\)](#)).

Of the 8 bills, [Facchini and Friedrich \(2011\)](#) classified 3 as major immigration legislations. The first major legislation was the *1990 Immigration Act* or *IMMACT*. Among other provisions, the *IMMACT* raised the annual cap for legal permanent residents to 675,000 from 530,000 ([Leiden and Neal \(1990\)](#)). This was followed by the *Illegal Immigration Reform and Immigrant Responsibility Act* of 1996. The bill's provisions included expedited removal procedure of undocumented immigrants and criminal penalty of those immigrants who re-enter or attempt to re-enter the U.S. within a specific time period after removal ([Grable \(1997\)](#)).

Of the bills post 9/11, a major bill was the *Border Protection, Anti-terrorism and Illegal Immigration Control Act* of 2005. One of the provisions of the bill, sponsored by Wisconsin republican Jim Sensenbrenner, was the construction of a 700 miles long fence along the US-Mexico border. Further, the bill required the federal government to take custody of locally detained illegal immigrants. The bill also would have made illegal presence in the United States an aggravated felony instead of a civil offence ([Fetzer \(2006\)](#)). Although the bill was passed in the house, it was later repealed by the Senate.

---

<sup>2</sup>ICPSR code is an identification number assigned to members of the Congress

Table 1: Final passage votes on immigration issues in the House of Representatives 1990-2006

Congress	Date	Bill	Topic	Keyword	Direction
101	3.10.1990	H.R.4300*	Immigration	The 1990 Immigration Act (IMMACT)	Pro
104	21.3.1996	H.R.2202*	Illegal Immigration	Illegal Immigration Reform and Immigrant Responsibility Act	Con
105	24.9.1998	H.R.3736	Skilled Immigration	Temporary Access to Skilled Workers and H-1B Nonimmigrant Program Improvement Act	Pro
109	10.2.2005	H.R.418	Illegal Immigration	Real ID Act	Con
109	16.12.2005	H.R.4437*	Illegal Immigration	Border Protection, Anti-terrorism and Illegal Immigration Control	Con
109	14.9.2006	H.R.6061	Illegal Immigration	Secure Fence Act	Con
109	21.9.2006	H.R.6094	Illegal Immigration	Community Protection Act of 2006	Con
109	21.9.2006	H.R.6095	Illegal Immigration	Immigration Law Enforcement Act of 2006	Con

This table is adapted from [Facchini and Friedrich \(2011\)](#). Major legislations on immigration during this period are marked using an asterisk (\*).

### 3 Methodology

#### 3.1 Immigration Corpus

All speeches are first cleaned by removing punctuation and upper case. The speeches are then lemmatized to reduce words to their base forms while maintaining the context. This is accomplished by using NLTK WordNet lemmatizer ([Bird et al. \(2009\)](#)).

From the congressional records, speeches containing the words *immigration* or *refugee* or related words are identified to form the “immigration corpus” ([Li and Hills \(2019\)](#)). The immigration corpus is what is referred to throughout the rest of the paper as a collection of all immigration related speeches made on the floor in the Congress. All procedural phrases are removed from the immigration speeches by using the list of procedural bigrams provide by [Gentzkow et al. \(2019\)](#). Following [Gentzkow et al. \(2019\)](#), all speakers’ last names and names of states and months are removed from the speeches. The paper excludes speeches where the speaker is identified by office and not by name (such as ‘chairman’). This yields a total of 24,351 immigration related speeches with speeches from 1098 distinct speakers. The corpus contains a total of 43,564,960 words and the average number of words per speech is 1789.

#### 3.2 Valence and Concreteness

For each speech the paper computes the valence (or sentiment) and concreteness (or proxy for social distance) ratings. For the former, the paper uses valence norms proposed by

Warriner et al. (2013). Warriner et al. (2013) provides valence norms for approximately 14,000 words, each rated on a scale of 1 to 9.<sup>3</sup> The valence rating of a word refers to the pleasant emotion conveyed by a word, with the rating increasing as it moves from unhappy to happy. Valence of a speech was computed by taking the mean valence rating of all words in the article. For concreteness, the paper uses concreteness norms proposed by Brysbaert et al. (2014). Brysbaert et al. (2014) offers the concreteness ratings of nearly 37,000 English words, each rated on a scale of 1 to 5, where 1 is very abstract and 5 is very concrete.<sup>4</sup>

### 3.3 Topic Model

The paper uses a Latent Dirichlet Allocation or LDA model (Blei et al. (2003)) to extract the different topics discussed in the immigration corpus. The LDA model assumes that there is a fixed number of topics or patterns which explains the structure of a corpus. LDA is a Bayesian model in which each document of a corpus is modelled as a finite mixture over an underlying set of topics.<sup>5</sup> The LDA model reduces each document in a corpus to a “bag-of-words” which is a representation of text that describes the number of occurrences of each word in the document, ignoring the sequence in which the words appear. The model assumes that a set of documents can be explained by a fixed number of underlying patterns or topics. Each document in the corpus is a distribution over these topics and each topic is a distribution over the words which form the corpus vocabulary.

For interpreting the topics extracted from the immigration corpus the paper examines the most relevant words pertinent to each topic. For this the paper uses the *relevance* measure proposed by Sievert and Shirley (2014). Let  $V$  denote the number of terms in the vocabulary of the immigration corpus,  $w$  be a word  $\in 1, \dots, V$ ,  $\phi_{kw}$  be the probability that  $w$  belongs to topic  $k \in 1, \dots, K$  and  $p_w$  be the marginal probability of word  $w$  in the immigration corpus, then the relevance of word  $w$  to topic  $k$  is defined as:

$$r(w, k|\lambda) = \lambda \log(\phi_{kw}) + (1 - \lambda) \log\left(\frac{\phi_{kw}}{p_w}\right) \quad (1)$$

where  $\lambda$  ( $0 \leq \lambda \leq 1$ ) is the relative weight parameter given to the probability that word  $w$  is assigned to topic  $k$  relative to the *lift* or  $\frac{\phi_{kw}}{p_w}$ .<sup>6</sup>

The paper aims to track trends in topics over different years, similar to Li et al.

---

<sup>3</sup>Each word rating was computed by taking the average rating of 20 participants.

<sup>4</sup>The concreteness rating of each word was computed by taking the average rating across 30 participants.

<sup>5</sup>The LDA model was implemented using *gensim* package in Python.

<sup>6</sup>The *lift* of a word (Taddy (2012)) is the ratio of a word’s probability within a topic to the word’s marginal probability of being in the corpus. The lift measure is used to rank terms within topics. This paper calculated the top relevant words for each topic by trying different values of  $k$ , between 5 to 15 and different values of  $\lambda$  between 0.5 to 1, to find the most contextually appropriate topics and keywords, presented in Table 5

(2018). This was done to understand the key topics associated with immigration over time. All speeches that occurred in the same year were considered as 1 document. For each document the paper calculates the probability that document  $d$  belongs to topic  $k$  as:

$$\phi_d(k) = \frac{|w \in d : \text{topic}(w) = k|}{|d|} \quad (2)$$

where the numerator is the number of words in document  $d$  that are assigned to topic  $k$  and the denominator is the total number of words in document  $d$ .

## 4 Results

The results are presented in 3 sections. Section 4.1 presents descriptive results examining the variation of immigrant speech valence and concreteness over time, by party and by region. Section 4.2 examines if linguistic measures of immigrant attitude can predict voting behaviour on immigration bills in the House of Representatives. Lastly, section 4.3 presents the results obtained from applying a topic modelling algorithm to the immigration corpus.

### 4.1 Descriptive Results

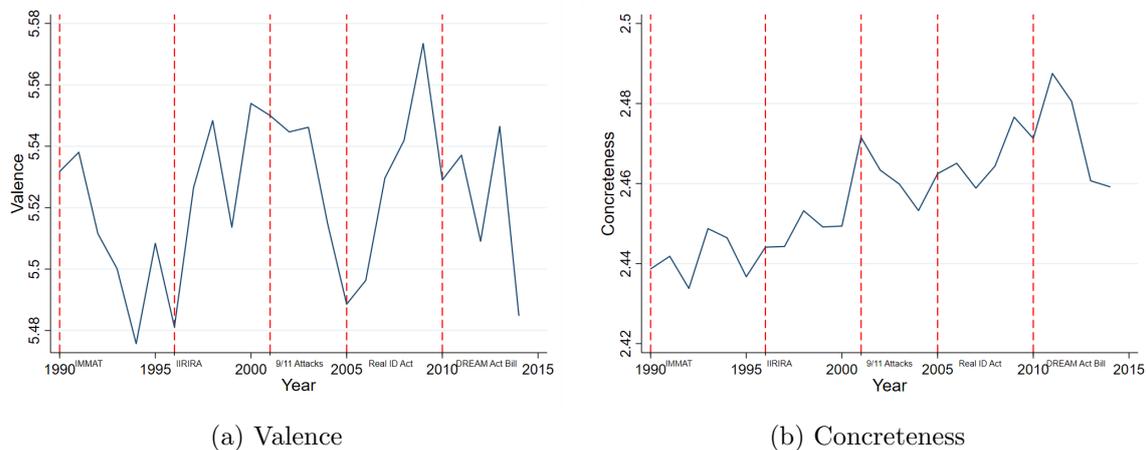


Figure 1: Variation in valence and concreteness of immigrant speeches over time

This section begins by looking at the variation in valence and concreteness measures of congressional speeches (in both chambers of the Congress) over time. Figures 1 (a) and (b) plots the mean valence and concreteness of immigration speeches over time, respectively. Figure 1 (a) shows a small rise in the valence measure following the Immigration Act (IMMACT) of 1990. The bill, which was signed into law by President George H. W. Bush, raised the annual cap on immigrant admissions, a positive step towards liberalising

immigration policy. The amount of attention paid to immigration policy varies by year. The number of votes on immigration measure (including votes on amendments and procedure) declined during the 103rd congress (1993-1994) and rose in the 104th Congress (1995-1996) (Fennelly et al. (2015)). The Illegal Immigration Reform and Immigrant Responsibility Act (IIRARA) of 1996, signed by President Clinton on 30th September 1996, aimed primarily at curbing the flow of illegal immigration into the United States (Grable (1997)). This is expressed by a steep drop in the valence measure leading up to the introduction of IIRARA. This is followed by an increase in valence measure leading to the House passing the Temporary Access to Skilled Workers and H-1B Nonimmigrant Program Improvement Act of 1998, temporarily increasing the annual number of H1B visas<sup>7</sup>.

There was an increase in valence towards late 1990s leading to the Congress enabling around 400,000 illegal immigrants, who had been deemed ineligible for earlier amnesties, to obtain green cards in December 2000 (Schuck (2005)). Further, in March 2001, 150,000 undocumented Salvadorans acquired temporary legal status. However, such positive immigrant measures came to an end with the attack on the World Trade Centre on September 11, 2001. There was a drastic decline in the valence measure after 2001. Following the attacks, there was a major modification of immigration policies in the US. Immigrant policy debates became centred on national security issues (Rodriguez (2008); Hing (2006); Schuck (2005)). The post 9/11 era saw the creation of the Department of Homeland Security in 2002, with the primary goals of preventing terrorist attacks and actively controlling the U.S. borders.

Immigration votes on national security and border concerns reached its peak in 2005-06 (Fennelly et al. (2015)). These votes included the Real ID Act of 2005, which, among other provisions, tightened the eligibility criteria for asylum seekers (Cianciarulo (2006)). This period also included other restrictive bills, in terms of liberalising Immigration policy, such as Border Protection, Anti-terrorism and Illegal Immigration Control Act of 2005, Secure Fence Act of 2005 and Immigration Law Enforcement Act of 2006. The valence measure improved after this under the Democratic control of Congress (both House and Senate) from 2007-2011. Around this time, there were several attempts to pass different versions of the Development, Relief, and Education for Alien Minors Act or DREAM Act, aiming to grant legal status to undocumented individuals who came to the U.S. as minors. It came closest to being passed in 2010 when the bill was passed in the House, but fell of short of the votes required to proceed in the Senate. Further, there was an increase in the valence measure following President Obama's announcement of the Deferred Action for Childhood Arrivals (DACA) in 2012, by executive Branch memorandum. The DACA program entitled individuals who had travelled to the U.S. as minors to apply for deferred action from deportation.

---

<sup>7</sup>The H1B visa allows U.S. organizations to hire skilled foreign workers for specialised jobs.

The concreteness measure, which is a proxy for social distance, has overall increased during the period under consideration. This is consistent with Hills et al. (2016) which finds that there has been an increase in concreteness in American English over time. Although, Figure 1 (b) shows a sharp fall in the measure following the 9/11 attacks in 2001.

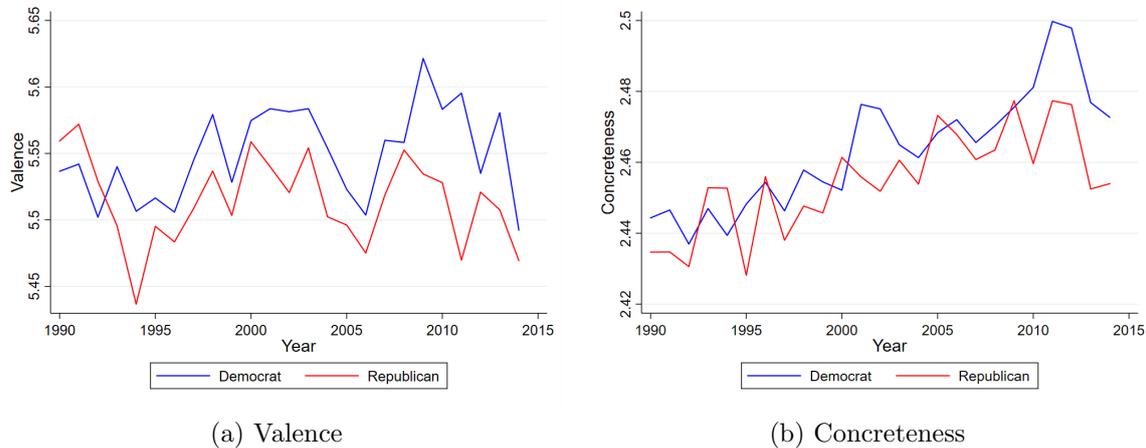
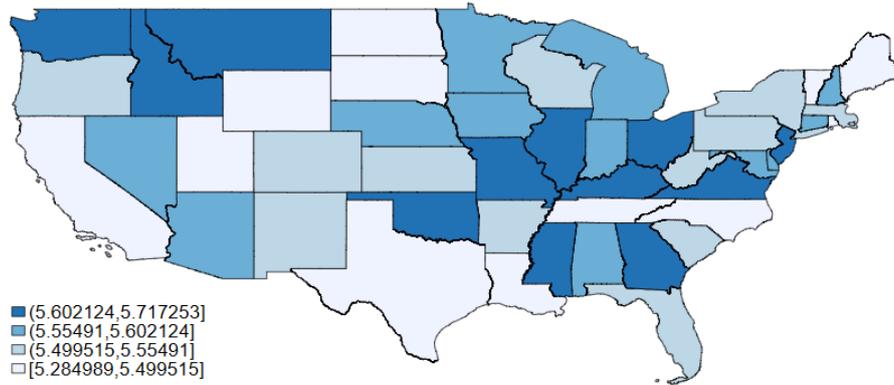


Figure 2: Variation in valence and concreteness of immigrant speeches of political parties over time

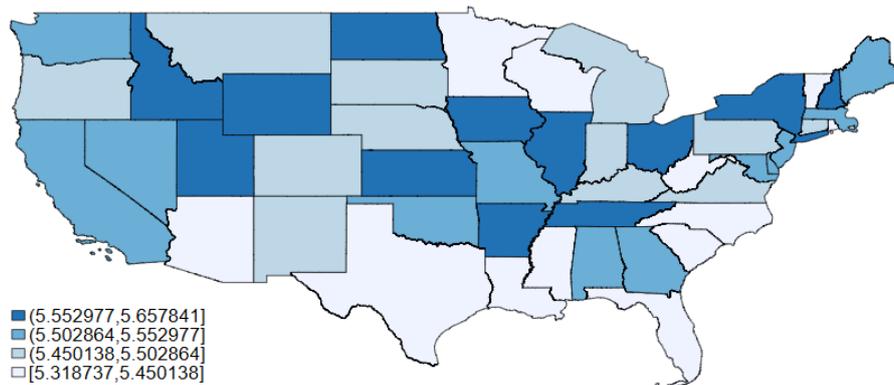
Next the speakers are matched with data on individual characteristics. There were 1026 speakers, from both chambers of the Congress, for whom biographical data was available. These were grouped according to their party affiliation to examine the variation of immigrant valence and concreteness over time, by party. There were 510 Republicans, 512 Democrat and 4 Independent legislators in the dataset. Figure 2 plots the results. Figure 2 (a) show that the valence measure for Democrats and Independents (grouped together and labelled as Democrat) has been higher than that for Republicans since early 1990s. The concreteness measure (Figure 2 (b)) has been higher for Democrats and Independents (grouped together and labelled as Democrat) than Republicans for most of the period under consideration. This suggests polarised attitude of the Congress members towards immigrants, with Democrats showing more pro-immigrant attitude.

Furthermore, the paper examines valence and concreteness across the different states in the US. Figure 3 shows the mean valence of immigration speeches across states in year 1990 (Figure 3 (a)) and 2006 (Figure 3 (b)). The lighter shades represent states with lower valence or immigrant sentiment. States with the lowest immigrant sentiment in 1990 were the southern states of Texas, Louisiana, Tennessee and North Carolina. Other low sentiment states were California, Utah and Wyoming in the west and North Dakota and South Dakota in the mid-west. Figure 3 (b) shows an increase in the number of southern states with low immigrant sentiment. In 2006 the southern states of Texas, Louisiana, Mississippi, Florida, South Carolina and North Carolina, showed the lowest average valence measures. The anti-immigrant attitude in the south is consistent with Fetzer (2006)

who found that people from the southern states in the U.S. oppose immigration owing to their conservative political nature. Figure A.1 shows the mean concreteness ratings of immigration speeches across states in 1990 and 2006. The variation across states is less systematic compared to the valence measure.



(a) Speech valence in 1990



(b) Speech Valence in 2006

Figure 3: Valence of immigration speeches across different states in the U.S. in (a) 1990 and (b) 2006

Figure 4 shows the average valence or sentiment of immigration related speeches over time in the different regions in the US. The Census Bureau of the United States divides the country into 4 regions - North-east, Midwest, South and West.<sup>8</sup> While there are clear overlaps, the north-eastern states appear to have had the most positive or highest immigrant sentiment over time. This is because of the strong democratic support in the North-eastern states, especially since the early 1990s (Harris (2014)). The southern states on the other hand have displayed a lower speech sentiment over time. Figure 4 (b) shows the sentiment for just the north-eastern and southern states for clearer comparison.

<sup>8</sup>The North-eastern States are Connecticut, Maine, New Hampshire, Massachusetts, New Jersey, New York, Pennsylvania, Rhode Island, Vermont. The Mid-western states are Illinois, Indiana, Iowa, Kansas, Michigan, Missouri, Minnesota, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin. The southern States include Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Oklahoma, North Carolina, South Carolina, Tennessee, Texas, Virginia, West Virginia. Lastly, the Western States are Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming.

Variation of concreteness measure over time, by region, was not as clear as the sentiment measure and is presented in Figure A.2.

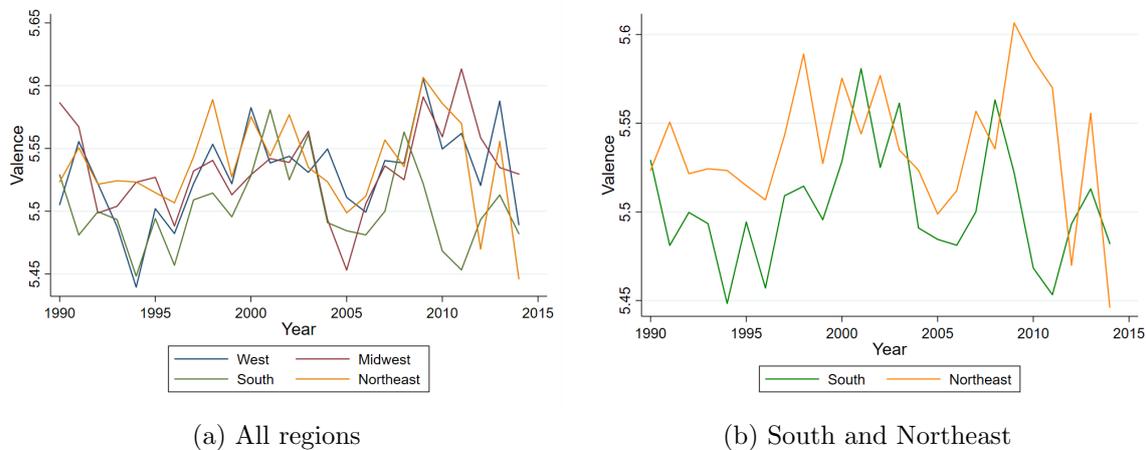


Figure 4: Variation in valence of immigration speeches over regions in the US. (a) shows valence over time for all regions (b) shows valence over time for the southern and north-eastern states.

## 4.2 Voting Behaviour

Having evaluated the variation in lexical attitude towards immigrants of congress members over time, by party and by region, the paper next studies their voting behaviour. The paper seeks to examine if the lexical measures computed for each speaker can predict their voting behaviour on bills related to immigrant issues. With regards to voting behaviour, the paper makes the following predictions:

*Prediction 1: Speakers with higher speech valence or greater pro-immigrant sentiment are more likely to vote in favour of bills liberalising immigration.*

Immigrant speech sentiment is predicted to be a powerful predictor of future voting behaviour on immigration bills. This result is expected to hold even after controlling for the speaker’s individual characteristics, party affiliation and congressional district characteristics. This is because speech valence captures the inherent opinion that congress members have about immigrant issues, above and beyond their party’s position and their district’s demographic and socio-economics features. Further, due to the high frequency of speeches made and the length of these speeches, a measure computed from speeches is more informative of the speaker’s voting intentions than other measures of immigrant attitude.

*Prediction 2: Speech valence is hypothesised to have strong predictive powers for voting behaviour even after limiting the dataset to include only republican majority states or states with high foreign born population.*

Members of the southern and western states in the US, owing to their conservative

political ideologies (Fetzer (2006)) have been observed to oppose immigration. In such states, where the sentiment towards immigrants is low (Figure 4), valence measure of speeches is expected to be a predictor of voting behaviour of congress members. Moreover, prediction 2 posits that in states which is majorly opposed to immigration, the sentiment towards immigrant is more polarised and hence the valence measure is expected to be a good predictor of voting behaviour.

Hawley (2011) observes that in the US, higher the local foreign born population in the state, greater the partisan influence on support for restricting immigration.<sup>9</sup> Accordingly, it is hypothesised that in such polarised states, the polarised sentiment towards immigrants is a good predictor of congress member’s support towards pro-immigration bills.

*Prediction 3: Concreteness measure of immigration speeches is not a good predictor of voting behaviour of congress members.*

Concreteness measure of immigration speeches has been observed to grow over time during the period considered (Figure 1). This implies that with the growth in immigrant population over time in the US, the perceived social distance towards immigrants has declined over time.<sup>10</sup> Further, due to the unsystematic variation of concreteness values across regions (Appendix B), concreteness is hypothesised to not be a good predictor of voting behaviour on immigrant issues.

The list of bills provided in Table 1 is used to examine the relationship between speaker speech sentiment and voting behaviour on immigration bills in the House of Representatives. The paper uses the probit model presented in equation (1) to evaluate the effect of immigrant sentiment on speaker’s voting behaviour.

$$Prob(Vote_{idt} = 1) = \Phi(\gamma_1 Valence_{idt} + \gamma_2 Z_i + \gamma_3 Y_{dt} + \delta_s + Post_t + \delta_s \times Post_t) \quad (1)$$

where the binary dependent variable  $Vote_{idt}$  takes the value 1 if the speaker  $i$  from district  $d$  votes in favour of liberalizing immigration at time  $t$ .  $\phi()$  is the cumulative distribution function. The key explanatory variable  $Valence_{idt}$  is the average sentiment or valence of all immigrant related speeches delivered by the speaker a year leading up to the vote. For example, if a vote occurs in March 1996, the value of  $Valence_{idt}$  will be the average valence value of all the immigration speeches delivered by the speaker between March 1995 and February 1996.  $Z_i$  accounts for the speaker  $i$ 's individual characteristics.  $Y_{dt}$  includes district level characteristics of district  $d$  in year  $t$ . The dummy variable  $Post_t$

---

<sup>9</sup>This is consistent with the *group threat theory* (LeVine and Campbell (1972); Blalock (1967); Blumer (1958)) which posits that competition for limited resources lead to conflict among different demographic groups i.e. larger the size of the out-group, greater the threat perceived by the in-group.

<sup>10</sup>This is consistent with *contact theory* (Allport (1954)) which suggests that biases towards outgroups are lessened though more interaction. This could also be because over time there has been an increase in concreteness in American English (Hills et al. (2016)).

takes the value 1 if the vote occurred after the 9/11 attacks and 0 otherwise.  $\delta_s$  are state dummies. The model also accounts for state specific shocks post 9/11 attacks by including  $\delta_s \times Post_t$  interaction terms.

The results of equation (1) are presented in table 3. The regression uses standardized values of the valence variable. The table reports the average marginal effects. Column 1 shows that an increase in 1 standard deviation in the valence or sentiment of the speaker’s speeches increases the probability of a pro-immigration vote by 5 percentage points, significant at the 5% level. Column 2 also accounts for the speakers’ characteristics such as whether the speaker is non-white or female and the speaker’s age. The valence measure still remains significant at the 5% level. Further, it is observed that being non-white significantly (at the 1% level) increases the probability of a pro vote. The variables age and female, though positive and significant in column 2, become insignificant on including additional control variables in column 3.

Table 2: District level variables

Variable	Definition
Foreign Born	The share of foreign born people in the total population in the district
Foreign Born Growth	The increase in the foreign born population compared to last period
African American	The share of African American people in the total population in the district
Hispanic	The share of people with Hispanic origin in the total population in the district
Democratic Vote Share	The Democratic share of the votes at the last House election.
Unemployment	The share of unemployed people in the total labour force in the district.
Highly Skilled ratio	The percentage of the population in the district over 16 who are employed in executive, administrative, managerial and professional speciality occupations
Unskilled ratio	The percentage of the population in the district over 25 years of age with less than 4 years of High School
Farm worker	The share of farm workers in the total labour force in the district.

The data for the district level variables used in the paper are obtained from [Facchini and Friedrich \(2011\)](#).

Columns 3, along with including the dummy variables Democrat (which equals 1 if the speaker is from a democratic party) and No military service (which equals 1 if the speaker has never performed any military service), also includes district level characteristics. The district level variables included along with their definition are provided in Table 2. Further, column 4 includes the contributions from labour and corporate political action committees (PACs), expressed as a share of total PAC contributions, as control variables.

This is included in the regressions because interest groups have been observed to shape migration policy (Facchini et al. (2011)).

Columns 3 and 4 of table 3 show that despite controlling for speaker specific characteristics, district level variables and interest group contributions, speaker speech sentiment remains a significant positive predictor of voting behaviour. Column 4 shows that an increase in 1 standard deviation in speaker speech valence increase the probability of a pro immigration votes by 2 percentage points, the impact significant at the 5% level. Further, among the control variables, it was observed that being a democrat increases the congress member's probability of a pro vote. Further, the share of unemployed and the share of farm workers has a significant negative effect on the probability of a pro vote. Also, the skill level of the population significantly impact voting behaviour. An increase in the share of highly skilled individuals in the district has a significant positive effect on pro immigration vote, consistent with the results of Facchini and Friedrich (2011) and Milner and Tingley (2011). However, contradictory to Facchini and Friedrich (2011), this paper also finds that an increase in the share of highly unskilled individuals in the district has a positive effect on the legislator's probability of voting in favour of liberalising immigration. Column 5 replicates column 4 by only considering the major votes as highlighted in Table 1. Speaker speech sentiment remains a positive and significant (at the 10% level) predictor of pro immigration vote.

*Result 1: Consistent with prediction 1, speaker speech sentiment is found to be a significant predictor of speaker's voting behaviour, even after controlling for speaker's individual characteristics, party affiliation and the district's socio-economic features.*

Furthermore, the paper evaluates the predictive power of speaker speech sentiment across different geographic specifications. Table 4 looks at the the predictive power of the valence measure in the southern, south-western and western states. Representatives from these regions have been observed to oppose policies liberalising immigration (Fetzer (2006)). The table also looks at the voting behaviour of representatives in high immigration states i.e. states with a large share of foreign born individuals.<sup>11</sup> Speaker speech valence is a positive and significant predictor of voting in favour of liberalising immigration in the south, south-west, west and the high immigration states.

In the southern states, an increase in 1 standard deviation in speaker speech valence significantly (at the 5% level) increase the probability of a pro vote by 3 percentage points. The effect is even stronger in the southwestern and western states. An increase in 1 standard deviation in speaker speech valence significantly (at the 5% level) increases the probability of a pro vote by 4 percentage points in the south-west and by 5 percentage points in the west. In the high immigration states an increase in 1 standard deviation in

---

<sup>11</sup>In the current dataset the high immigration states were Arizona, California, Connecticut, Florida, Georgia, Illinois, Massachusetts, Maryland, New Jersey, Nevada, New York, Rhode Island, Texas, Virginia and Washington.

Table 3: Impact of speech sentiment on Immigration votes

	All bills				Major bills
	(1)	(2)	(3)	(4)	(5)
Valence	0.0476*** (0.0177)	0.0283** (0.0140)	0.0167* (0.0097)	0.0242** (0.0100)	0.0284* (0.0167)
Non-white		0.4714*** (0.0492)	0.1262*** (0.0475)	0.1575*** (0.0457)	0.1685** (0.0762)
Age		0.0037** (0.0018)	-0.0001 (0.0013)	-0.0013 (0.0014)	-0.0049** (0.0019)
Female		0.0957** (0.0483)	0.0125 (0.0335)	0.0059 (0.0359)	-0.0286 (0.0544)
Democrat			0.2564*** (0.0407)	0.2883*** (0.0608)	0.2939*** (0.0854)
No Military Service			-0.0404 (0.0309)	-0.0447 (0.0347)	-0.0562 (0.0448)
Foreign Born			-0.0851 (0.2306)	-0.4402* (0.2354)	-0.2257 (0.4578)
Foreign Born Growth			-0.0442** (0.0206)	-0.0239 (0.0187)	-0.0490** (0.0233)
African American			0.1714 (0.1539)	0.1180 (0.1512)	-0.1603 (0.2166)
Hispanic			-0.1403 (0.1633)	0.1088 (0.1802)	-0.0265 (0.2232)
Democratic Vote Share			0.1980* (0.1032)	0.1565 (0.1047)	0.1192 (0.1484)
Unemployment			-1.2480 (1.0413)	-2.4244** (1.1118)	-2.1876 (1.5581)
Highly Skilled Ratio			0.6117** (0.2805)	1.0202*** (0.3377)	0.8588 (0.5407)
Unskilled Ratio			1.3138*** (0.3236)	1.7441*** (0.3760)	1.9765*** (0.4972)
Farm worker			-1.6431** (0.7112)	-2.5525*** (0.7411)	-2.1445** (0.9935)
Labour PAC				-0.0900 (0.1534)	-0.0067 (0.2179)
Corporate PAC				-0.1671 (0.1139)	-0.1817 (0.1573)
Post 9/11	-0.0784 (0.1511)	-0.1018 (0.1271)	-0.1129 (0.0940)	-0.1395 (0.1442)	0.1137 (0.2035)
<i>N</i>	1137	1082	1082	844	316

Standard errors, clustered at the speaker level, in parentheses

All regressions control for state fixed effects and state interacted with Post 9/11 dummy fixed effects.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

valence significantly (at the 10% level) increases the probability of a pro vote by 2 percentage points. Further, being a non-white significantly increases the legislator’s probability of a pro vote in the south and in the high immigration states. Being a democrat has a significant effect on the probability of a pro vote across all 4 specifications considered.

Among the district level variables, the share of foreign born individuals has a positive effect on the probability of a pro vote in the southern states. This is consistent with the findings of [Facchini and Friedrich \(2011\)](#) and [Milner and Tingley \(2011\)](#). However, this effect is reversed in the south-western and western states. Consistent with table 3, a large share of high skilled individuals has a positive effect on a pro vote in the west and in the high immigration states. Also, consistent with table 3, an increase in unskilled individuals has a significant positive effect on pro immigration voting behaviour. The share of farm workers is negatively associated with pro voting behaviour.

*Result 2: Even after limiting the dataset to just consider the republican majority states and high immigration states, speaker speech sentiment remains a significant predictor of speaker voting behaviour on immigration bills.*

The predictive power of concreteness measure of speeches was tested for all votes and for votes in southern, south-western, western and high immigration states. The results are presented in tables A.3 and A.4. For all regions, as well as for the selected geographic specifications, concreteness was not found to be a good predictor of voting behaviour.

*Results 3: Consistent with Prediction 3, concreteness was observed not to be a good predictor of speaker voting behaviour on immigration bills.*

### 4.3 Immigrant Topics

This section presents the results from applying a Latent Dirichlet Allocation (LDA) model on the immigration corpus. This topic modelling algorithm helps to understand the key topics for immigrant legislation, discussed in the U.S. Congress. Table 5 presents 10 relevant words describing each of the 8 topics that were identified. Each topic was labelled by careful consideration of the keywords<sup>12</sup>. The topics reflect the possible categories of immigration legislation. The 8 topics are immigrant benefits or assistance, border security, national security, defense, educational policy, health care, policy amendment and U.S. Government support.<sup>13</sup>

Next the paper examines the distribution of these topics over time. Figure 5 plots the time variation of the 8 topics. The distribution of topics over time expresses the variation

---

<sup>12</sup>To maximize the dataset for training the LDA model the paper uses all speeches containing immigration, refugee and related words even those where the speaker is identified by office and not by name. The LDA model was applied to 26,467 immigration related speeches in both chambers of the Congress.

<sup>13</sup>The value of  $\lambda$  was set at 1 and the top 20 keywords were evaluated for each topic. Of the top 20 words the 10 most interpretive words (and excluding words with no semantic interpretation such as would, year, section, subsection) are presented in the paper. Experimenting with different values of  $\lambda$  between 0.5 to 1 yielded very similar results.

Table 4: Impact of speech sentiment on Immigration votes in different regions

	South	Southwest	West	High Immigration
	(1)	(2)	(3)	(4)
Valence	0.0331** (0.0156)	0.0419*** (0.0158)	0.0499** (0.0207)	0.0211* (0.0113)
Non-white	0.1464*** (0.0359)	-0.0201 (0.0597)	-0.0209 (0.0763)	0.0835* (0.0505)
Age	-0.0005 (0.0020)	-0.0021 (0.0022)	-0.0009 (0.0027)	0.0001 (0.0015)
Female	0.1265* (0.0674)	-0.0041 (0.0641)	-0.0691 (0.0721)	0.0214 (0.0403)
Democrat	0.2062*** (0.0532)	0.3338*** (0.0741)	0.3452*** (0.1159)	0.2391*** (0.0484)
No Military Service	0.0093 (0.0458)	-0.0071 (0.0526)	0.0015 (0.0575)	-0.0339 (0.0351)
Foreign Born	1.2495*** (0.3710)	-1.0623*** (0.3401)	-1.9541*** (0.4693)	-0.0794 (0.2506)
Foreign Born Growth	0.0399** (0.0175)	-0.1120*** (0.0406)	-0.1200*** (0.0462)	-0.0802*** (0.0274)
African American	-0.0332 (0.1992)	-0.1835 (0.3402)	-0.3169 (0.4909)	0.1845 (0.1885)
Hispanic	-0.3181* (0.1894)	0.1793 (0.2496)	0.4698 (0.4147)	-0.0691 (0.1901)
Democratic Vote Share	0.0152 (0.1154)	0.2849 (0.2137)	0.2943 (0.3122)	0.1784 (0.1192)
Unemployment	1.3779 (1.2407)	-1.7474 (1.3157)	-2.9285 (1.9384)	-1.1945 (1.3109)
Highly Skilled Ratio	-0.7411* (0.3906)	0.9259 (0.5781)	2.4481*** (0.7374)	0.6858** (0.3310)
Unskilled Ratio	0.0351 (0.3271)	1.7023** (0.6787)	3.6025*** (1.1269)	1.4614*** (0.3996)
Farm worker	-0.9956 (1.2290)	-2.5328*** (0.9347)	-4.6462*** (1.3900)	-2.0781** (0.8940)
Post 9/11	-0.5713*** (0.0955)	0.0147 (0.0962)	-0.1451 (0.2772)	-0.0417 (0.2177)
<i>N</i>	369	325	260	734

Standard errors, clustered at the speaker level, in parentheses

All regressions control for state fixed effects and state interacted with Post 9/11 dummy fixed effects.

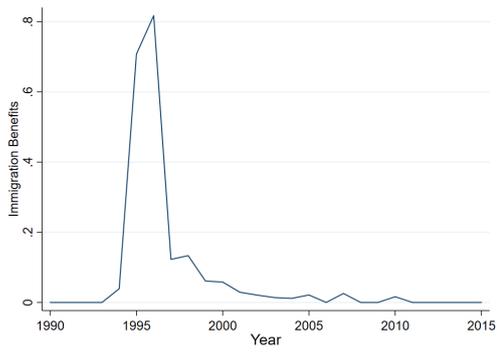
\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 5: Immigrant Topics and keywords

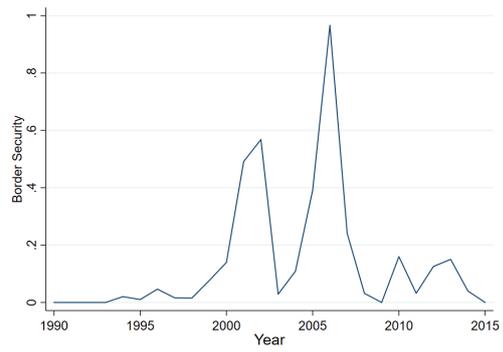
	Topic	Keywords
1	Immigrant Benefits	child, program, amendment, law, fund, assistance, public, government, health, provision
2	Border security	security, act, law, border, national, department, amendment, bill, work, right
3	National Security	security, national, intelligence, report, agency, service, government, federal, program, fund
4	Defense	law, fund, national, defense, service, bill, fund, department, American, act
5	Educational Policy	program, school, child, federal, fund, educational, education, law, assistance, amount
6	Health care	health, care, act, fund, provided, law, amount, plan, available, security
7	Policy Amendment	act, program, service, fund, amount, amended, law, fiscal, available, section
8	U.S. Government Support	refugee, law, war, government, American, support, need, right, world, program

in concern over different immigrant issues and is generally reflected by the nature of the bills introduced during the time. Figure 5 (a) shows that discussion on benefits and rights of immigrants peaked in 1995-96, consistent with the findings of Fennelly et al. (2015). Figure 5 (b) shows that border security concerns peaked first in 2002 leading to the formation of the Department of Homeland Security in 2002, post the 9/11 attacks. The second peak in Border security legislation was in 2005-2006 with bills such as Border Protection, Anti-terrorism and Illegal Immigration Control Act of 2005, Secure Fence Act of 2005 and Immigration Law Enforcement Act of 2006. Similarly, Figure 5 (c) shows a peak in national security legislation around 2005. For defense legislation, Figure 5 (d) shows a peak in 2000 with the introduction of National Defense Authorization Act for Fiscal Year 2001. This is followed by another peak between 2010-2014. This period marked the military involvement of the U.S. in the Syrian Civil War.

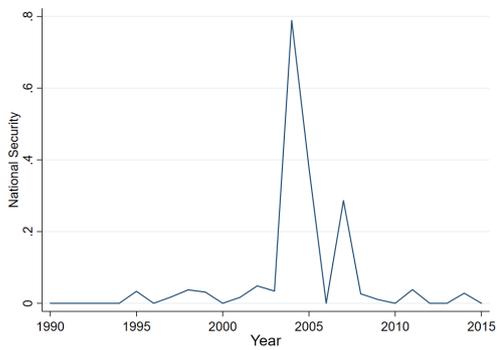
Educational policy topics peaked in 1994 with the Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations Act, which involved immigrant provisions (Fennelly et al. (2015)) and had scattered attention after that till 2001, following which it received little to no attention. Health care topics peaked with the Children’s Health Insurance Program Reauthorization Act of 2009, which included immigrant provisions (Fennelly et al. (2015)). Not surprisingly, discussion over policy amendments in Congress have been scattered all over the period under consideration. Finally, discussions related to U.S. government support or aid was maximum between 1990-1994, a period which included the First Gulf War.



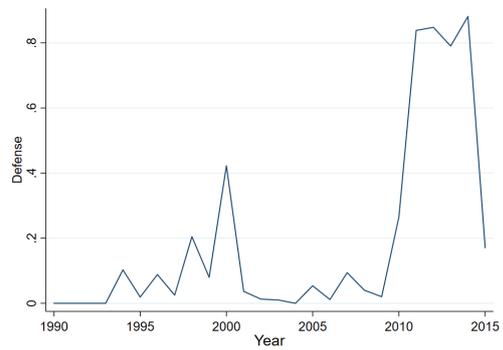
(a) Immigration Benefits



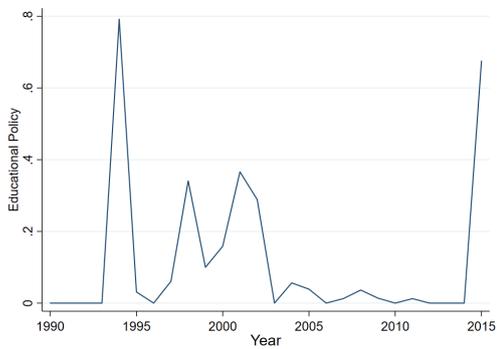
(b) Border Security



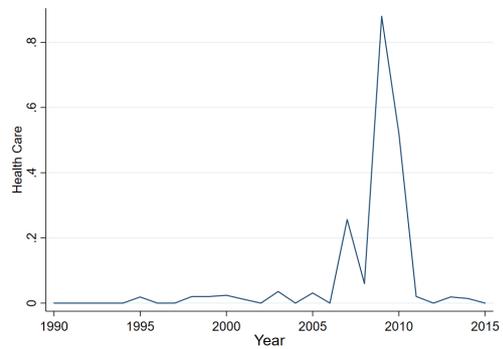
(c) National Security



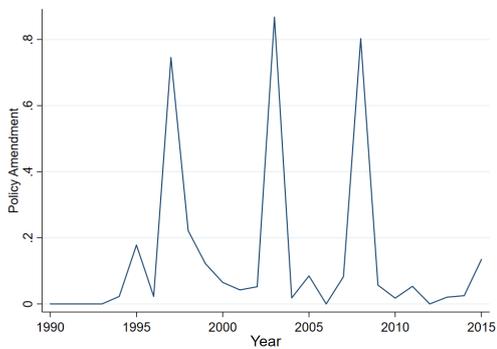
(d) Defense



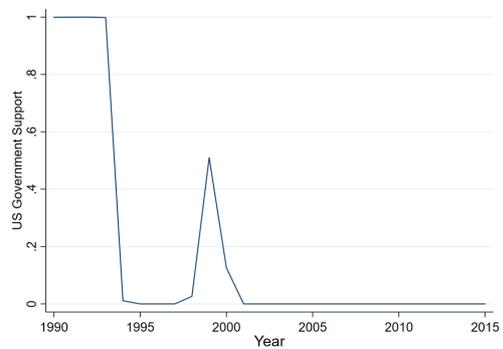
(e) Educational Policy



(f) Health Care



(g) Policy Amendment



(h) US Government Support

Figure 5: Immigrant Topics over Time

## 5 Conclusion

This paper creates an “immigration corpus” which contains 24,351 congressional speeches containing *immigration*, *refugee* and related words, delivered on the floor of congress between 1990 and 2015. The corpus is used to build two distinct lexical measures of attitude towards immigrants - immigrant *sentiment* or valence and language *concreteness*. Out of these two, immigrant sentiment showed systematic variation over time, consistent with the history of immigrant outcomes in the US, with a marked decline following the September 2011 terrorist attacks in the U.S. and the subsequent introduction of several anti-immigration bills. The sentiment measure was strictly polarised with democrats displaying a consistently higher sentiment than republicans in the period considered. The sentiment measure also displayed orderly variation over regions, with the conservative states in the southern and western regions showing consistently low immigrant sentiment. The concreteness measure, barring fluctuations around major immigrant events such as the 9/11 attacks, displayed an overall increasing trend during the period considered.

The paper also computed speaker specific values of immigrant speech sentiment to show that speech sentiment is a significant positive predictor of speakers’ voting behaviour on immigration related votes. This result holds even after controlling for speaker characteristics, party affiliation and district level socio-economic variables.

Further, the paper applies a Latent Dirichlet Allocation or LDA topic modelling algorithm to the immigration corpus to identify the different immigration related topics discussed in the U.S. congress and also study the variation of topic distribution over time. The topic distribution shows a hike in topics like border security and national security following the 9/11 attacks and a decline of topics like educational policies and government aid.

Congressional speeches delivered on issues related to immigration are vital for predicting immigrant outcomes in the US. A lexical measure of immigrant attitude developed using speeches is superior for predicting immigrant outcomes compared to other measures like past voting behaviour of speakers on immigration bills. This is because the former provides a continuous measure thus capturing the *intensity* of the speaker’s attitude compared to the latter which is a binary measure. Further, the U.S. congress members play an important role in shaping legislation related to immigrant issues by voting either in favour of or against bills which liberalise or restrict immigration. A lexical measure of immigrant attitude is useful for predicting the speaker’s voting behaviour.

Overall, the study of speeches delivered by congress members either during Congressional sessions or during campaign trails and other public appearances is important to form a measure of the politician’s attitude towards immigration or other key issues. Future research, can use a similar methodology as used in the paper, to analyse politicians’ attitude on other crucial issues such as gender, climate change and taxation and use it to

predict the politician's voting behaviour on bills related to such issues.

## References

- Allport, G. W. (1954). *The nature of prejudice*. New York: Doubleday.
- Bird, S., E. Klein, and E. Loper (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. O'Reilly Media.
- Blalock, H. M. (1967). *Toward a Theory of Minority-Group Relations*, Volume 325. New York: Wiley.
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan), 993–1022.
- Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific sociological review* 1(1), 3–7.
- Brysbaert, M., A. B. Warriner, and V. Kuperman (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46(3), 904–911.
- Cianciarulo, M. S. (2006). Terrorism and asylum seekers: Why the real ID act is a false promise. *Harvard Journal on Legislation* 43, 101.
- Facchini, G. and M. Friedrich (2011). What drives U.S. immigration policy? Evidence from congressional roll call votes. *Journal of Public Economics* 95(7-8), 734–743.
- Facchini, G., A. M. Mayda, and P. Mishra (2011). Do interest groups affect us immigration policy? *Journal of International Economics* 85(1), 114 – 128.
- Fennelly, K., K. Pearson, and S. Hackett (2015). The US Congressional Immigration Agenda: Partisan Politics, Policy Stalemate and Political Posturing. *Journal of Ethnic and Migration Studies* 41(9), 1412–1432.
- Fetzer, J. S. (2006). Why Did House Members Vote for H.R. 4437? *International Migration Review* 40(3), 698–706.
- Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica* 87(4), 1307–1340.
- Goldin, C. (1994). The political economy of immigration restriction in the United States, 1890 to 1921. In *The regulated economy: A historical approach to political economy*, pp. 223–258. University of Chicago Press.

- Gonzalez, J. G. and N. Kamdar (2000). Do Not Give Me Your Tired, Your Poor! Determinants of Legislator Voting on Immigration Issues. *Eastern Economic Journal* 26(2), 127 – 143.
- Grable, D. M. (1997). Personhood under the due process clause: A constitutional analysis of the illegal immigration reform and immigrant responsibility act of 1996. *Cornell L. Rev.* 83, 820.
- Harris, W. (2014). Essays on the Political Geography of The United States of America.
- Hawley, G. (2011). Political Threat and Immigration: Party Identification, Demographic Context, and Immigration Policy Preference. *Social Science Quarterly* 92(2), 404–422.
- Hills, T. T., J. S. Adelman, and T. Noguchi (2016). Attention economies, information crowding, and language change. *Big Data in Cognitive Science*, 270.
- Hing, B. O. (2006). Misusing immigration policies in the name of homeland security. *CR: The New Centennial Review* 6(1), 195–224.
- Jensen, J., S. Naidu, E. Kaplan, L. Wilse-Samson, D. Gergen, M. Zuckerman, and A. Spiraling (2012). Political polarization and the dynamics of political language: Evidence from 130 years of partisan speech. *Brookings Papers on Economic Activity*, 1–81.
- Leiden, W. R. and D. L. Neal (1990). Highlights of the US Immigration Act of 1990.
- LeVine, R. A. and D. T. Campbell (1972). *Ethnocentrism: Theories of conflict, ethnic attitudes, and group behavior*. John Wiley & Sons.
- Li, Y. and T. Hills (2019). Quantifying Historical Change in Patterns of Immigrant Sentiment.
- Li, Y., T. Hills, and R. Hertwig (2018, November). A brief history of risk. *PsyArXiv*. <https://doi.org/10.31234/osf.io/8vqhm>.
- Maass, A., D. Salvi, L. Arcuri, and G. R. Semin (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of personality and social psychology* 57(6), 981.
- Mastro, D., R. Tukachinsky, E. Behm-morawitz, E. Blecha, D. Mastro, R. Tukachinsky, E. Behm-morawitz, E. Blecha, D. Mastro, R. Tukachinsky, E. Behm-morawitz, and E. Blecha (2014). News Coverage of Immigration : The Influence of Exposure to Linguistic Bias in the News on Consumer’s Racial/Ethnic Cognitions. 3373.
- Milner, H. V. and D. H. Tingley (2011). The economic and political influences on different dimensions of United States immigration policy. *Available at SSRN 2182086*.

- Rice, D., J. H. Rhodes, and T. Nteta (2019). Racial bias in legal language. *Research & Politics* 6(2), 2053168019848930.
- Rodriguez, R. M. (2008). (Dis)unity and Diversity in Post-9/11 America. *Sociological Forum* 23(2), 379–389.
- Schuck, P. H. (2005). Immigrants’ Political and Legal Incorporation in the United States after 9/11: Two Steps Forward, One Step Back. *SSRN Electronic Journal*.
- Sherman, A., D. Trisi, C. Stone, S. Gonazales, and S. Parrott (2019). Immigrants Contribute Greatly to U.S. Economy, Despite Administration’s “Public Charge” Rule Rationale. , Centre on Budget and Policy Priorities. Accessed 27 February 2020 at <https://www.cbpp.org/sites/default/files/atoms/files/8-15-19pov.pdf>.
- Sievert, C. and K. Shirley (2014). Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pp. 63–70.
- Snefjella, B. and V. Kuperman (2015). Concreteness and psychological distance in natural language use. *Psychological Science* 26(9), 1449–1460. PMID: 26239108.
- Taddy, M. (2012). On estimation and selection for topic models. In *Artificial Intelligence and Statistics*, pp. 1184–1193.
- Trope, Y. and N. Liberman (2010). Construal-level theory of psychological distance. *Psychological review* 117(2), 440.
- Turner, J. (1985). Social Categorization and Self-Concept: A Social Cognitive Theory of Group Behavior. In E. Lawler (Ed.), *Advances in Group Process: Theory and Research*, pp. 77–121. Greenwich, Connecticut: JAI Press.
- United Nations (2017). International Migration Report 2017: Highlights. ST/ESA/SER.A/404, United Nations, Department of Economic and Social Affairs, Population Division. Accessed 27 February 2020 at [https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2017\\_Highlights.pdf](https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2017_Highlights.pdf).
- United Nations (2019). International Migration 2019: Report. ST/ESA/SER.A/438, United Nations, Department of Economic and Social Affairs, Population Division. Accessed 27 February 2020 at [https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/InternationalMigration2019\\_Report.pdf](https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/InternationalMigration2019_Report.pdf).

Warriner, A. B., V. Kuperman, and M. Brysbaert (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45 VN - r(4), 1191–1207.

# A Appendix

## A Additional Tables

Table A.1: Characteristics of Voters

Characteristic	Number of Speakers
Republican	203
Democrat	219
Independent	1
Female	48
Non-White	77
No Military Service	304
Total Speakers	445

Table A.2: Summary Statistics of Voting Data

Variable	Mean	Std. Dev.	Min.	Max.	N
Vote	0.391	0.488	0	1	1265
Valence	5.526	0.194	4.635	6.262	1265
Age	56.211	9.481	30	81	1213
Foreign Born	0.112	0.108	0.002	0.585	1265
Foreign Born growth	0.507	0.810	-0.718	5.254	1265
African American	0.123	0.161	0.001	0.801	1265
Hispanic	0.118	0.151	0.003	0.745	1264
Democratic Vote Share	0.51	0.259	0	1	1260
Unemployment	0.059	0.025	0.018	0.204	1264
High Skilled Ratio	0.316	0.087	0.116	0.582	1264
Unskilled ratio	0.22	0.093	0.044	0.623	1265
Farm Worker	0.019	0.023	0	0.204	1264
Labour PAC	0.195	0.191	0	0.888	1010
Corporate PAC	0.369	0.15	0	1	1010

Table A.3: Impact of speech concreteness on Immigration votes

	All bills				Major bills
	(1)	(2)	(3)	(4)	(5)
Concreteness	-0.0044 (0.0172)	0.0084 (0.0162)	0.0064 (0.0120)	0.0130 (0.0125)	0.0244 (0.0188)
Non-white		0.4805*** (0.0498)	0.1299*** (0.0477)	0.1663*** (0.0470)	0.1740** (0.0795)
Age		0.0037** (0.0019)	-0.0001 (0.0013)	-0.0011 (0.0014)	-0.0046** (0.0020)
Female		0.1001** (0.0486)	0.0119 (0.0336)	0.0067 (0.0371)	-0.0322 (0.0559)
Democrat			0.2562*** (0.0411)	0.2845*** (0.0599)	0.2968*** (0.0852)
No Military Service			-0.0365 (0.0309)	-0.0363 (0.0338)	-0.0447 (0.0457)
Foreign Born			-0.0614 (0.2309)	-0.3835 (0.2372)	-0.2031 (0.4596)
Foreign Born Growth			-0.0434** (0.0211)	-0.0207 (0.0194)	-0.0453* (0.0246)
African American			0.1636 (0.1556)	0.1074 (0.1555)	-0.1952 (0.2190)
Hispanic			-0.1565 (0.1619)	0.0763 (0.1775)	-0.0580 (0.2217)
Democratic Vote Share			0.1993* (0.1042)	0.1642 (0.1057)	0.1067 (0.1449)
Unemployment			-1.1203 (1.0496)	-2.2365** (1.1219)	-1.8928 (1.5861)
Highly skilled Ratio			0.6235** (0.2846)	1.0326*** (0.3517)	0.9385* (0.5669)
Unskilled Ratio			1.3087*** (0.3263)	1.7287*** (0.3798)	1.9529*** (0.5045)
Farm worker			-1.6471** (0.7024)	-2.5004*** (0.7321)	-2.0986** (0.9878)
Labour Pac				-0.0805 (0.1515)	0.0091 (0.2261)
Corporate PAC				-0.1418 (0.1093)	-0.1485 (0.1565)
Post 9/11	-0.1175 (0.1590)	-0.1294 (0.1311)	-0.1308 (0.0965)	-0.1694 (0.1465)	0.0600 (0.2097)
<i>N</i>	1137	1082	1082	844	316

Standard errors in parentheses

All regressions control for state fixed effects and state interacted with Post 9/11 dummy fixed effects.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A.4: Impact of speech concreteness on Immigration votes in different regions

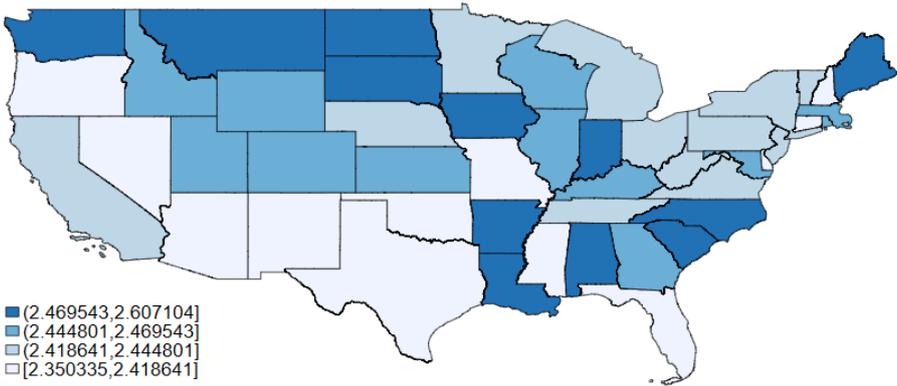
	South	Southwest	West	High Immigration
	(1)	(2)	(3)	(4)
Concreteness	0.0113 (0.0169)	0.0093 (0.0226)	-0.0007 (0.0280)	0.0214 (0.0152)
Non-white	0.1572*** (0.0338)	-0.0105 (0.0599)	-0.0051 (0.0758)	0.0872* (0.0512)
Age	-0.0003 (0.0021)	-0.0017 (0.0021)	-0.0009 (0.0027)	0.0003 (0.0015)
Female	0.1105 (0.0693)	0.0112 (0.0630)	-0.0353 (0.0691)	0.0197 (0.0405)
Democrat	0.1999*** (0.0504)	0.3297*** (0.0721)	0.3316*** (0.1101)	0.2361*** (0.0494)
No Military Service	0.0212 (0.0451)	-0.0091 (0.0551)	-0.0089 (0.0616)	-0.0275 (0.0354)
Foreign Born	1.2688*** (0.3703)	-0.9215*** (0.3166)	-1.7137*** (0.4446)	-0.0463 (0.2483)
Foreign Born Growth	0.0500*** (0.0165)	-0.1195*** (0.0398)	-0.1252*** (0.0462)	-0.0817*** (0.0276)
African American	-0.0503 (0.2012)	-0.2020 (0.3521)	-0.3737 (0.5058)	0.1773 (0.1896)
Hispanic	-0.2971 (0.1900)	0.1410 (0.2535)	0.4187 (0.4327)	-0.0767 (0.1855)
Democratic Vote Share	0.0290 (0.1079)	0.2751 (0.2018)	0.3273 (0.3017)	0.1856 (0.1224)
Unemployment	1.7573 (1.2729)	-1.4930 (1.2786)	-2.2908 (1.8888)	-1.0070 (1.3029)
Highly Skilled Ratio	-0.8132** (0.3934)	0.9388* (0.5441)	2.2346*** (0.7128)	0.6885** (0.3363)
Unskilled Ratio	-0.1527 (0.3171)	1.6153** (0.6278)	3.1452*** (1.0886)	1.4216*** (0.3982)
Farm worker	-1.0110 (1.1748)	-2.2773*** (0.8558)	-4.1269*** (1.2882)	-2.0721** (0.8502)
Post 9/11	-0.5912*** (0.0960)	-0.0038 (0.0980)	-0.1020 (0.2401)	-0.0133 (0.1921)
<i>N</i>	369	325	260	734

Standard errors in parentheses

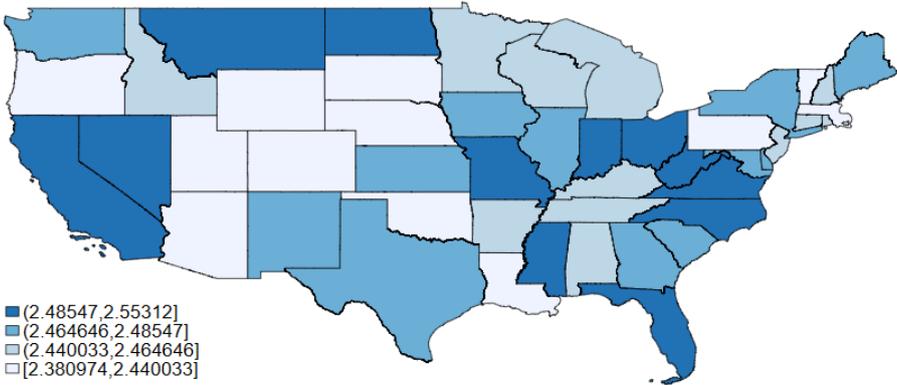
All regressions control for state fixed effects and state interacted with Post 9/11 dummy fixed effects.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## B Additional Figures

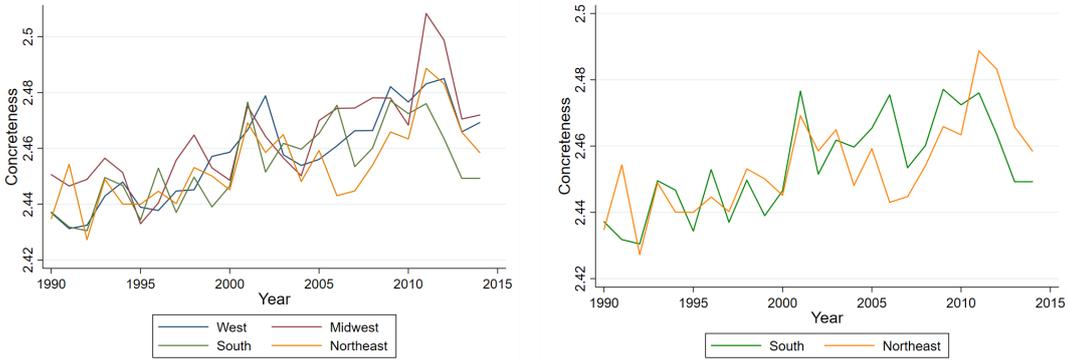


(a) Speech concreteness in 1990



(b) Speech concreteness in 2006

Figure A.1: Concreteness of immigration speeches across different states in the U.S. in (a) 1990 and (b) 2006



(a) All regions

(b) South and Northeast

Figure A.2: Variation in concreteness of immigration speeches over regions in the US. (a) shows concreteness over time for all regions (b) shows concreteness over time for the southern and north-eastern states.