ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

MASTER THESIS

Programme International Economics

THE RISE OF COVID-19: GOOGLE TRENDS PREDICTABILITY OF ANTI-ASIAN HATE CRIME

Name Student: Maring, Lukas

Student ID Number: 572698

Supervisor: Prof. Dr. Olivier Marie

Second Assessor: Marco Musumeci

Date Final Version: 30th April 2022

Abstract

The theoretical framework of this paper proposes the use of Google search popularity of racial slurs as a proxy for racial bias. Evidence has validated racial bias as the prerequisite for ethnic hate crime. By exploring Google Trends' capability to predict anti-Asian hate crime in the United States using data from 2004 to 2019, this research adds to the existing research on employing Google Trends to predict (hate) crime. Further, in order to support this paper's findings for predicting hate crime against Asians, Google Trends capability to predict hate crime against Hispanics is examined. The "Trump-effect" is briefly discussed in order to account for a structural break in anti-Hispanic hate crime believed to originate from Trump becoming the 45th president of the United States of America. The findings of this research concluded that Google Trends' data is a poor predictor of anti-Asian and -Hispanic hate crime. Future, research trying to predict hate crime will most likely find a better fitting model using more traditional methods of prediction.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

As of April 28th 2022, the COVID-19 pandemic, originating in Wuhan City, Hubei Province of China (WHO, 2020), officially left 509.53 million confirmed cases and has taken the lives of approximately 6.23 million worldwide (WHO, 2022). Furthermore, based upon March 17th 2021 figures, it has pressed governments to offer fiscal packages totalling $16 trillion (IMF, 2021). During the beginning of the pandemic, U.S. media coverage (Ismael & Measor, 2003; Nyamnjoh, 2010; Ahmed 2021) and the racist rhetoric of the Trump administration have labelled the coronavirus as the “Wuhan Virus” and “Chinese Virus” (Benjamin, 2020; Chiu, 2020; Fallows, 2020). This and the role of social media (Croucher et al., 2020; Darling-Hammond et al., 2020) have enforced racial bias against Asian Americans and individuals of Asian descent, in the United States. In response to the rise in violence against Asians, brought about through the perpetuation of racial bias against Asians, the United Nations released a statement on March 22nd 2021 in which Secretary-General António Guterres voiced alarm over the increased victimisation of members belonging to the Asian community since the beginning of the COVID-19 pandemic (United Nations Secretary-General, 2021). Stop AAPI Hate, a non-profit organisation which records hate and discrimination incidents towards Asian and Pacific Islanders in the U.S., reported over 10,152 incidents from March 19th 2020 till December 31st 2021 – according to their national report (Stop AAPI Hate, 2022). The report highlights the invasiveness of such incidents with 32.4% of hate and discrimination occurring in public, 26.9% in the work place and 10.3% in the proximity of private residence.

Shively (2005) expresses, in the National Institute of Justice report *Study of Literature and Legislation on Hate Crime in America, Final Report*, that the ability to forecast and predict a given region’s level of hate crime may help measure the realised or potential effect of criminal justice programmes and hate crime preventative policies. Successful crime prevention programmes can be cost-effective (Welsh et al., 2015) by mitigating added excessive burdens to victims’ general welfare (Miller et al., 1993; Barnes et al., 1994) and mental well-being (Cohen et al., 1998; Cornaglia et al., 2014; Benier, 2017), reducing the impact of hate crime on societal and physical segregation (Perry, 2009), and dampening the cost to society and the economy (Detotto et al., 2010; McCollister et al., 2010). However, the use of statistical analysis, algorithms or prediction models in “predictive policing” by law enforcement, with the idea to prevent crime, do not automatically result in optimal law enforcement as they may subvert law enforcement officers’ and agencies’ decision making and responsibility (Bennett et al., 2016; Williams et al., 2019).

Recent literature related to anti-Asian hate crime (Chen et al., 2020; Gover et al., 2020; Tessler et al., 2020; Wu et al., 2020; Zhang et al., 2021) belongs to the broader religious, ethnic and sexual orientation hate crime literature, such as anti-Jewish (Iganski, 2007; Kielinger, 2007; Klikauer, 2018; Mills, 2020), anti-Islamic (Kaplan, 2006; Disha et al., 2011; Awan et al., 2015, 2016; Borell, 2015; Gardell, 2015; Ivandic et al., 2019), anti-Hispanic (Cummings et al., 1997; Stacey et al., 2011; Light et al., 2015) and anti-LGBTQ+ (Meyer, 2010; Duncan et al., 2014; Mills, 2019; Kehoe, 2020).

Similar to Stephens-Davidowitz (2014), Google search popularity of racial slurs (Google Trends data) is utilised as a proxy for racial bias, the prerequisite for ethnic hate crime. By exploring how well Google Trends predicts anti-Asian hate crime in the United States using data from 2004 to 2019, this paper attempts to expand upon research on utilising Google Trends to predict crime and hate crime predicting literature [1]. Although much less literature exists for the former, the likes of Gamma et al. (2016) and Piña-García and Ramírez-Ramírez (2019) have utilised Google Trends in order to predict *meth*-related crimes and a variety of committed crimes in Mexico City, respectively. Findings within this paper, such as the disparity between Asians and Hispanics willingness to report hate crime, adds to prevailing hate crime predicting literature (Williams et al. 2019; Jendryke et al., 2021; Wang, 2021). Additionally, in order to help validate and add to this paper's findings for predicting hate crime against Asians, Google Trends capability to predict hate crime against Hispanics is examined. Comparable to Asians, Hispanics have been victims of hate crime throughout their history in the U.S., belong to the ethnic-minorities and often seen as not belonging in the United States (Cummings et al., 1997; Mindiola et al., 2009; Stacey et al., 2011; Gratton and Merchant, 2013; Light et al., 2015)[2].

During Donald Trump's presidential campaign and much of his presidency, the expression of racial prejudices by Americans in the public sphere rose at the expense of immigrants and ethnic minorities due to Trump's contentious rhetoric (Newman et al. 2020) – this has been coined by some as the "Trump effect" (Costello, 2016). Research by Müller et al., (2018) and Rushin et al. (2018) provides evidence of Trump's impact on hate crime provoked by his racially-biased commentary. As an extension, this paper addresses a structural

[1] The year 2020 is not covered in this paper as at the time of research, FBI hate crime data for 2020 had not been published.

[2] Blacks or African-Americans and Middle Easterners ethnic groups were not used to validate results. The former because of their history, culture and time spend within the United States potentially too dissimilar to that of Asians. The latter due to the unavailability of anti-Middle Easterners hate crime data between 2004 and 2019.

break in anti-Hispanic hate crime believed to be the consequences of the Trump effect, aimed at improving prediction performance.

This paper begins Section 2 which covers a brief historical overview of Asian discrimination in the U.S and introduces this paper's theoretical framework in. The components of the national- (time-series) and state-level (panel) datasets as well as data transformation are expounded upon in Section 3. Subsequently, the applied methodology is explained (Section 4) and main results presented (Section 5). Section 6 investigates the impact accounting for methodology bias and structural breaks has on prediction performance. In Section 7 results, possible limitations and how findings of this paper fair with relevant literature. Lastly, this paper is concluded in Section 8.

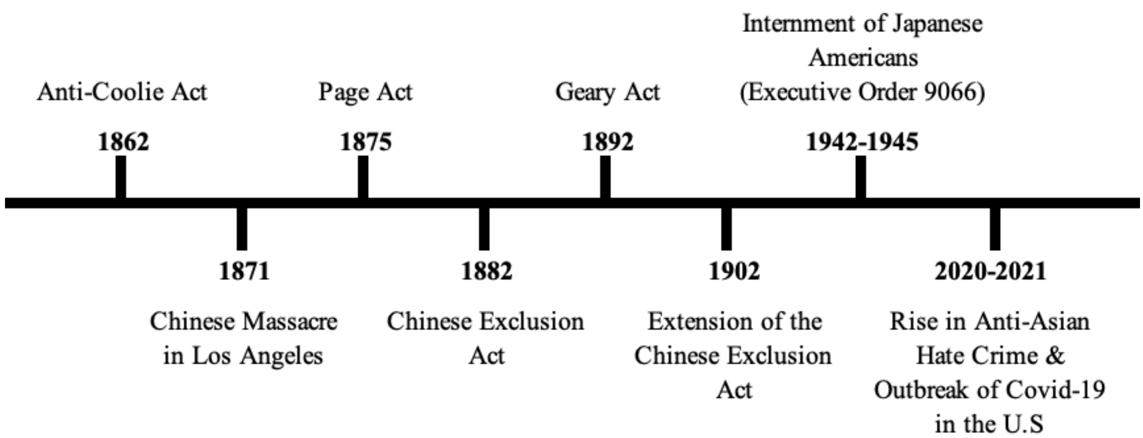
2 BACKGROUND & THEORY

A short historical summary of persecution and injustice faced by Asians in the United States is conveyed in Section 2.1, whilst this paper's theoretical framework is introduced in Section 2.2.

2.1 Brief History of Asian Discrimination

Historically, there has been much animosity in the United States towards the Asian community, resulting in both violent encounters and anti-Asian policies (Jung, 2005; Lee, 2007; Chang, 2009). The rise in anti-Asian hate crime during the COVID-19 outbreak is not an isolated event where Asians have been discriminated or violently victimised, as portrayed by Figure 2.1 – an incomplete list with examples of both violent and non-violent historical events in which Asians have been victimised by U.S. citizens or their government.

Figure 2.1: Timeline of Historical Anti-Asian Events in the U.S.



In 1860, *An Act to Prohibit the 'Coolie Trade' by American Citizens in American Vessels*, thereafter known as the *Anti-Coolie Act,* was passed by the 37th U.S. Congress which prohibited any vessel travelling to the U.S. to carry Chinese citizens "… to be disposed of, or sold, or transferred, for any term of years or for any time whatever" (U.S. Congress, 1860, p. 340)[3]. Although initially introduced in order to restrict and deter former slaveholders, no longer able to hold black slaves due to slavery abolishment, from importing or smuggling Asians into the country for economic gains, the *Anti-Coolie Act* paved the way for later discriminatory immigration laws – predominantly against Asian labourers (Jung, 2005). Thereafter, the *Page Act of 1875* (U.S. Congress, 1873) transferred power to U.S. consul-generals located at ports to deny any immigrants of Asian origin entry if it was determined that the immigrant's future employment was classed as "lewd" and "immoral". Subsequently, intensified systematic discrimination targeted Chinese through the *Chinese Exclusion Act* (U.S. Congress, 1881) by suspending the entry to the U.S., residency there in and citizenship thereof. In 1892, the U.S. Congress (1892) appended the 1882 *Chinese Exclusion Act* through the *Geary Act* which extended the regulating of persons of Chinese descent entering and residing in the country by ten years and made it compulsory for Chinese persons to, at all times, carry proof of legal immigration. The *Geary Act* empowered the federal government's ability to impose immigration controls and ultimately deportation and detention of illegal aliens. All anti-Chinese immigration laws were extended indefinitely following the extension of the *Chinese Exclusion Act* passed in the 1902 57th U.S. Congress (1902)[4].

Although previous anti-Chinese immigration laws were partially repealed via the 1943 *Magnuson Act* (U.S. Congress, 1943) in an attempt for the U.S. to seek favour with the Chinese (allies against the Japanese during World War II), the Japanese targeting *Executive Order 9066* (1942) was issued a year earlier in response to the military attack against Pearl Harbor by the Empire of Japan. The order enforced the physical removal and detention of all persons of Japanese heritage to fenced and guarded internment camps, consequently leading to an estimated combined property and net income loss of $4 billion for the victims of *Executive Order 9066* (National Archives, 2022).

[3] Violation of this legislation resulted in a financial fine of no more than two thousand U.S. dollars and a maximum of one year imprisonment.

[4] It is important to note that this extension was reverse following the 1943 Magnuson Act, Immigration and Nationality Act of 1952 and 1965.

The Chinese massacre of 1871 in Los Angeles, marked by Lee (2013) as the largest lynching in the United States, may not have been directly incited by judicial or political changes, nonetheless, highlights Asians painful and less than ideal inclusion and acceptance in the U.S. Almost a 150 years later, Asians have once again been targets of physical, emotional and verbal abuse, being associated with the outbreak of COVID-19, initially discovered in Wuhan, China. Over 2020 and 2021, a total of 10,905 incidents of hate against Asians were reported with harassment (66.9%), physical violence (16.2%) and avoidance and shunning (16.1%) being the top types of discrimination (Stop AAPI Hate, 2022).

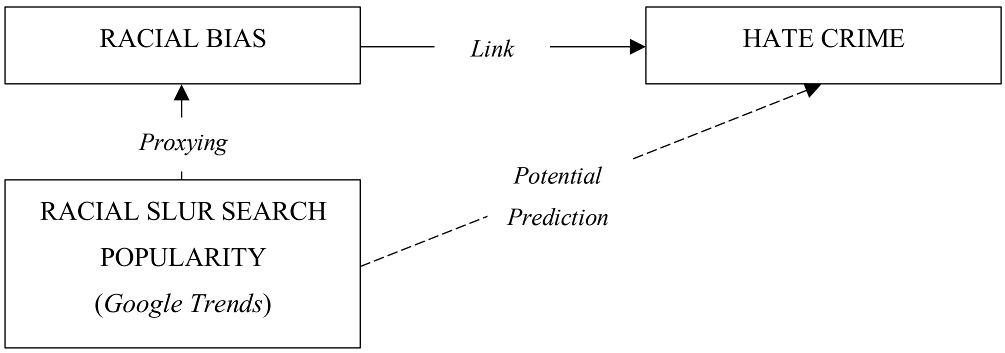
Although, the above mentioned events don't fully capture the long, complex and profound history of antagonism toward individuals of Asian origin, it provides a structured overview of systematic and non-governmental Asian discrimination in the U.S.. Illustrating that the recent rise in anti-Asian hate crime is not an isolated event provides ample motivation to explore models which can predict anti-Asian hate crime. With the capability to predict rises in hate crime comes the possibility for the government to implement preventative programmes before rises in anti-Asian hate crime comes to fruition – an early warning detection approach.

2.2 Theoretical Framework

The underlying theoretical framework of this paper, depicted in Figure 2.2, is dependent, firstly, on both racial bias as a prerequisite for the occurrence of ethnic hate crime and, secondly, on the ability of search popularity for racial slurs to proxy for racial bias. Looking closer at the definition of "hate crime" aids in validating racial bias as a precondition for hate crime. Powers et al. (2018) discuss definitions for hate crime, where "the key feature… is that it is motivated by an animus or prejudice against an entire group of people." Synonymously, the Hate Crime Statistics Act, which demands the collection of hate crime statistics from the government, defines hate crime as "…crimes that manifest evidence of prejudice based on race, gender and gender identity, religion, disability, sexual orientation, or ethnicity" (Hate Crime Statistics Act, 1990, 28 U.S.C. § 534). Although racial prejudice doesn't automatically lead to the committing of hate crime, it is clear that racial bias is fundamental for the existence of hate crimes.

Stephens-Davidowitz (2014) argues that Google Trends data is capable of capturing and proxying for racial bias through the search popularity of racial slurs. In his work, Stephens-Davidowitz argues that the use of Google Trends data (non-survey-based) may represent a more accurate picture of a society's racial bias than any conducted survey. Based upon work from Kreuter et al. (2008), Stephens-Davidowitz suggests that that Google users, of whom the majority are often alone when accessing the internet, are less likely to feel socially pressured to avoid expressing or searching socially distasteful theology than in social settings such as in-person conducted surveys. Research from Tourangeau et al. (2007), Berinsky (1999) and Kuklinski (1997) adds to the previous argument by proposing the potential underestimation of socially inappropriate mindsets, such as racial bias, from surveys.

Figure 2.2: Theoretical Framework



Greater racial bias has been linked to greater levels of hate crime (Williams et al., 2019; Müller et al., 2020; FBI, 2020). Google Trends data's potential to proxy for socially uncensored racial bias and link between racial bias and hate crime, provides argument for the opportunity to predict hate crime using online racial slur search popularity.

3 DATA

Monthly racial slur search popularity (Section 3.1), hate crime (Section 3.2), population and employment (Section 3.3) data has been sourced from various institutions such as Google, the Federal Bureau of Investigation (FBI), Bureau of Justice, United States Census Bureau (USCB) and Bureau of Labour Statistics (BLS) for the time period of 2004-2019 to use for both the national- and state-level datasets. Section 3.4 details the transformation applied to data in preparation for use in the main results.

The state-level dataset is constructed using monthly data collected by U.S. state. For certain U.S. states racial slur search popularity data is unavailable resulting in the exclusion of these states from the state-level results, as shown by Table 3.1. To resolve this issue for the time-series dataset, since the aggregation of the state-level dataset is used to produce the national-level (time-series) dataset, search popularity data is separately collected for the national-level – additional reasons for this approach are discussed in Section 3.1.

Table 3.1: List of Included U.S. States in State-level Results

U.S. State	Included	U.S. State	Included	U.S. State	Included
Alabama	X	Kentucky		North Dakota	
Alaska		Louisiana	X	Ohio	X
Arizona	X	Maine		Oklahoma	X
Arkansas		Maryland	X	Oregon	X
California	X	Massachusetts	X	Pennsylvania	X
Colorado	X	Michigan	X	Rhode Island	
Connecticut	X	Minnesota	X	South Carolina	X
Delaware		Mississippi		South Dakota	
District of Columbia*	X	Missouri	X	Tennessee	X
Florida	X	Montana		Texas	X
Georgia	X	Nebraska	X	Utah	X
Hawaii	X	Nevada	X	Vermont	
Idaho		New Hampshire		Virginia	X
Illinois	X	New Jersey	X	Washington	X
Indiana	X	New Mexico		West Virginia	
Iowa	X	New York	X	Wisconsin	X
Kansas	X	North Carolina	X	Wyoming	X

*- the District of Columbia is technically classed as a "federal district" and not a U.S. state.

3.1 Google Trends Data

3.1.1 Collection & Construction of Data

Search popularity of specific racial slurs is used as proxy variable for racial bias. This data is collected through publicly available Google Trends data. Collecting data for the search popularity of specific racial slurs, is rather time consuming because of the manner in which Google Trends normalises each search term's data dependent on the other search terms in the set of multiple search terms being retrieved. The search popularity scale lies between 0 and 100, representing no or extremely little popularity and highest search popularity in the given time period, respectively. Equation 3.1 expresses how the data for a given racial slur (r) in a given geographical area (s) in a given time period (t) is constructed:

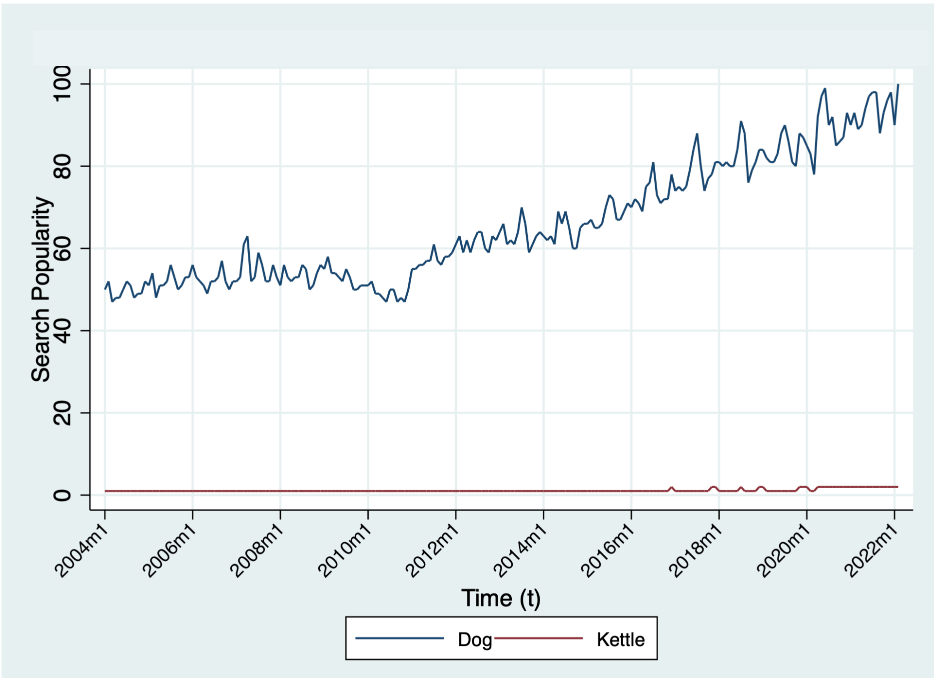
$$Term\ Search\ Popularity_{r,s,t} = \left[\frac{Number\ of\ Google\ Searches\ for\ Term_{r,s,t}}{Total\ number\ of\ Google\ Searches_{s,t}}\right] \qquad (3.1)$$

$$Normalised\ Term\ Search\ Popularity_{r,s,t} = 100 \times \left[\frac{Term\ Search\ Popularity_{r,s,t}}{Maximum\ Term\ Search\ Popularity_{s}}\right] \qquad (3.2)$$

Shown by Equation 3.2, after search popularity has been calculated for each racial slur, its corresponding *Term Search Popularity* value is normalised by Google Trends to fit on the scale from 0 to 100 based on the maximum *Term Search Popularity* value of the racial slur for which data is being collected – the maximum value is selected across time and terms. For example, in Figure 3.1 the maximum search popularity belongs to the term *dog* in January 2022. This is the *Maximum Term Search Popularity* by which all other data points for *dog*, as well as *kettle* are normalised.

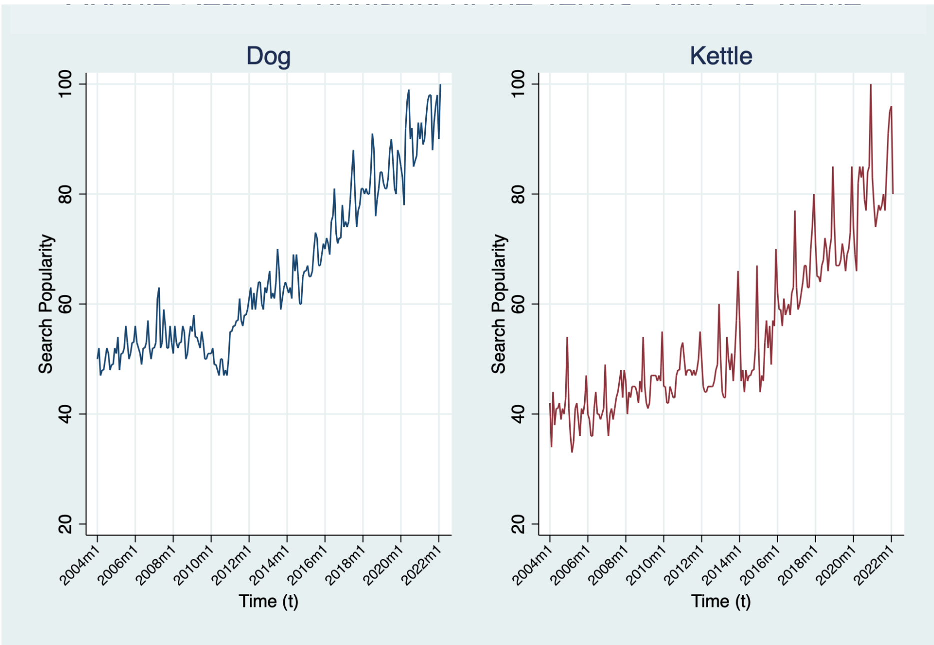
This means that when retrieving data for a set of terms consisting of more than two terms (a maximum of five terms) from Google Trends, it is possible that a term that is searched less has its values heavily transformed due to the much larger *Maximum Term Search Popularity* value of one of the other racial slurs within the set – so as shown in Figure 3.1 with the term *kettle*. Consequently, multiple term data collection leads to delivering little useful information when looking at much less searched racial slurs' search popularity as their variances are strongly pulled towards zero. This is best observed by comparing the search popularity variation across time for the terms *dog* and *kettle* in multiple and single data collection shown in Figure 3.1 and 3.2, respectively.

Figure 3.1: Search Popularity of *dog* vs. *kettle* using Multiple Term Data Collection



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

Figure 3.2: Search Popularity of *dog* vs. *kettle* using Single Term Data Collection



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

Therefore, to allow for better capturing of a term's search popularity relative to itself, Google Trends data is collected on a term-by-term basis for the national-level dataset, instead of in sets of multiple terms. However, using single term data collection for the state-level dataset, although more accurate than multiple term collection, would result in a computational nightmare as search popularity data can only be collected one U.S. state at a time[5].

Multiple term data collection raises two issues. The first problem arises from Google Trends' normalisation method pulling the variance of lesser searched terms in a set towards 0. Often rendering the data almost completely useless for the purpose of this paper. Furthermore, as data for 6 racial slurs is needed per ethnic group, it is necessary to create two racial slur retrieval sets for both anti-Asian and anti-Hispanic slurs. This introduces the second issue of racial slurs being normalised differently across sets since each set is normalised dependent on the *Maximum Term Search Popularity* of the set. This heterogeneity in data normalisation between racial slur sets would create errors in the data collection. Thus, to reduce the computational burden and accommodate for equal normalisation within anti-Asian and -

[5] The computation nightmare would involve separately downloading a .csv file for each U.S. state and each racial slur, ultimately, leading to manually downloading 600 files for search popularity of 12 racial slurs.

Hispanic slur sets, an "anchor term" is used to allow every racial slur's search popularity to be normalised based on a common exogenous *Maximum Term Search Popularity* value[6].

A successful anchor term has (i) little variation over time and (ii) isn't searched significantly more than the other terms it is compared to across all geographical areas of interest Fulfilling these requirements helps to reduce the racial slur term's loss of search popularity variation. A perfect anchor term doesn't exist because the search popularity of the term across time differs between U.S. states. However, the anchor terms "Lisbon" and "Indian Springs" for anti-Asian and anti-Hispanic slurs, respectively, generally fulfil the previously mentioned requirements and are therefore used as anchor terms in this method.

One caveat of this data collection method does not resolve for the state-level setting, however, is the fact that each racial slur's search popularity variation continues to be partially dependent on another term rather than based on its own search popularity. It will be necessary to take this into consideration when analysing state-level results.

3.1.2 Racial Slurs

In the previous section we've discussed how Google Trends data is constructed and retrieved. The following section focuses on the selection of the specific racial slurs used in the Google Trends data collection.[7]

Table 3.2: Complete List of Anti-Asian Slurs

Racial Slur	**Originally Targeted Ethnicity**	**Excluded**	**Reason for Exclusion**
Chink(s)	Chinese		-
Gook(s)	Korean		-
Flip(s)	Filipinos	X	*flip(s)* is mainly searched in combination with sports and technological devices.
Hapa(s)	Islanders		-
Jap(s)	Japanese		-
Locust(s)	Hong Kongese	X	*locust(s)* is mainly searched in combination with the animal, food for animals and a locations.
Paki(s) (Pakistani)	Pakistanis		-
Slant(s)	Chinese	X	*slant(s)* is mainly searched in combination with geometry.

[6] An example of the used anchor terms can be found in the Appendix Figures A.1 and A.2. They show the search popularity of anti-Asian and anti-Hispanic slurs with their anchor terms for Michigan and California, respectively.
[7] For both anti-Asian and -Hispanic slurs, data is collected for the singular and plural e.g. "gook" and "gooks" are treated as two separate words.

Slope(s)	Vietnamese	X	*slope(s)* is mainly searched in combination with geometry.
China doll(s)	Chinese	X	*china doll(s)* is mainly searched in combination with glazed porcelain dolls.
Chinaman(s)	Chinese		-
Ching chong(s)	Asian		-
Dragon lady(s)	Asian		-
Buddha head(s)	Chinese		-
Squint eye(s)	Chinese		-
Yellow belly(s)	Chinese	X	*yellow belly(s)* is mainly searched in combination with animals e.g. *yellow belly snake* or *yellow belly turtle.*
Chinee(s)	Chinese		-
Chinkie(s)/Chinky(s)	Chinese		-
Chow(s)	Chinese	X	*chow(s)* is mainly searched in combination with the dog breed or food.
Quang(s)	Vietnamese	X	*quang(s)* is mainly searched in combination with famous Vietnamese individuals or Vietnamese food.
Yellow bastard(s)	Chinese	X	*yellow bastard(s)*, the singular is mainly searched in combination with a New York City based brand and the plural has no Google Trends data.
Coolie	Indians		-

If a racial slur was searched disproportionately for a different intent or topic than to express racial bias, it was excluded. Table 3.2 shows the complete list of collected racial slurs with reasons for excluding certain terms. The list draws from research by Croom (2018) which discusses U.S. anti-Asian stereotypes and slurs. A few slurs have also been picked out of work from Hughes (2008) which offers additional terms used by the English-speaking world to degrade ethnic-minorities[8].

The list of anti-Hispanic slur is drawn from research by Croom (2014) which reviews the contextual beginnings of the Mexican-American demeaning slurs – the list of terms is shown in Table 3.3.

Table 3.3: Complete List of Anti-Hispanic Slurs

Racial Slur	**Originally Targeted Ethnicity**	**Exclduded**	**Reason for Exclusion**
Wetback(s)	Mexicans		-
Beaner(s)	Mexicans		-
Bronc(s)	Mexicans		-
Chopa(s)	Mexicans		-
Chopita(s)	Mexicans		-

[8] Croom (2018) and Hughes (2009) provide background on the origin of the listed anti-Asian slurs.

Greaser(s)	Mexicans	X	*greaser(s)* is mainly searched in combination with 1950s subculture and fancy dress.
Jagger(s)	Mexicans	X	*jagger(s)* is mainly searched in combination with the famous musician "Mick Jagger".
Pepper belly(s)	Mexicans		-
Roach coach(s)	Mexicans	X	*roach coach(s)* is mainly searched based for its meaning as "food truck" rather than a racial slur.
Taco bender(s)	Mexicans		-
Spic(s)	Mexicans		-

National Google Trends data is retrieved for all of the non-excluded anti-Asian and -Hispanic slurs as listed in Table 3.2 and 3.3. However, due to the lack of data for a number of U.S. states, Google Trends data is only retrieved for racial slurs with available data on the state-level – these terms are shown in Table 3.4.

Table 3.4: List of Racial Slurs for State-level

Anti-Asian Slurs	Anti-Hispanic Slurs
Gook(s)	Wetback(s)
Ching Chong*	Beaner(s)
Japs**	Spic(s)
Paki(s)	

* - the plural "ching chongs" was dropped due to insufficient available state-level data.

** - the singular "jap" was excluded due to it mainly being searched with the intention of searching for the country Japan or Japanese topics.

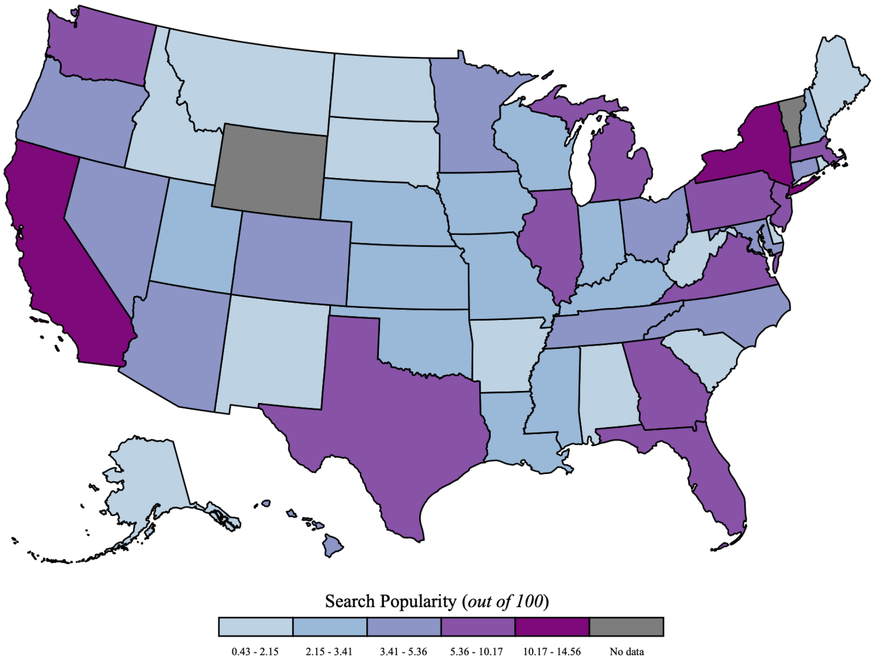
To capture the potential non-linear predictive relationship between racial slurs and hate crime, racial slur search popularities are squared and added to the dataset. For example, both *gook* and *gooks²* are included.

Figure 3.3 and 3.4 show the average monthly search popularity, based on Google Trends data, of anti-Asian and -Hispanic slurs, respectively, over the years 2004-2019. The average monthly search popularity per U.S. state is the summation of all anti-Asian or anti-Hispanic racial slur search popularity values, for a given U.S. state across 2004-2019, and divided by the total number of months in that time period[9]. This allows for a general overview of the monthly search popularity across different racial slurs for each U.S. state.

[9] Not all U.S. states have search popularity data available for all six slurs, therefore, the average search popularity is calculated on at least a minimum of one racial slur. Only Wyoming and Vermont have no available anti-Asian search popularity data for any of the six racial slurs – this is also seen in their omittance in Figure 3.3.

Figure 3.3: Average Monthly Search Popularity of Anti-Asian Slurs

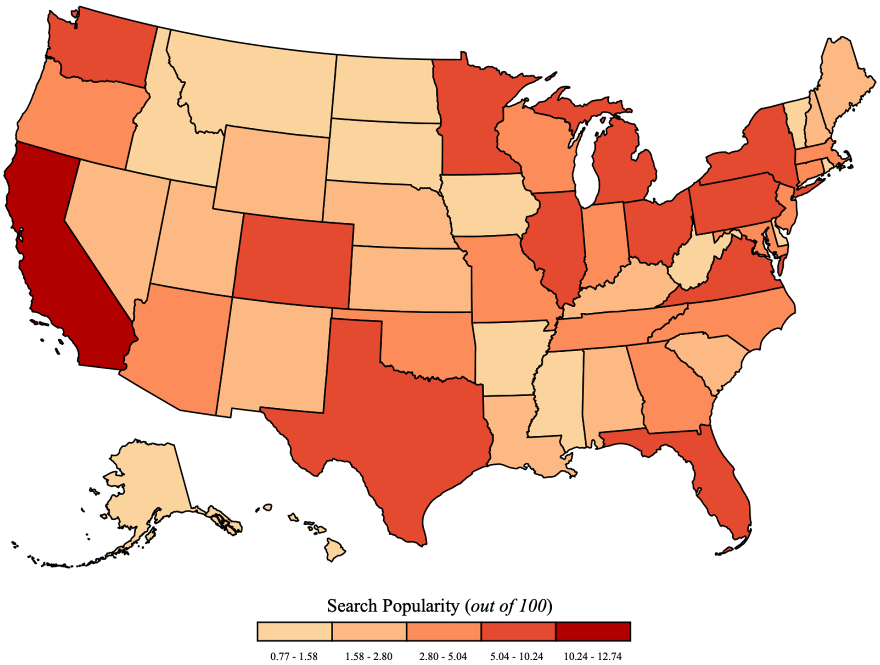
(State-level, 2004-2019)



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

Figure 3.4: Average Monthly Search Popularity of Anti-Hispanic Slurs

(State-level, 2004-2019)



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

Interestingly, California, compared to all other states, has high levels of average search popularity for anti-Asian and -Hispanic slurs – second highest anti-Asian slur search popularity, behind New York, and highest anti-Hispanic slur search popularity. Initially, it would seem plausible that these high levels could be related to California's Asian and Hispanic population making up on average, over 2004-2019, 51.3% of its total population – the highest ratio of all U.S. states. However, this explanation fails with respects to the U.S. states New Mexico and Hawaii, the second and third highest average Asian and Hispanic population ratio of 48.1% and 47.6% respectively, who's average racial slur search popularity falls in the bottom 50%. One theory which may hold more traction is the link between U.S. states' average total populations and average racial slur search popularity, with Table 3.5 showing the 8 largest populated states and corresponding average racial slur search popularity. It is conceivable that higher populated U.S. states attract more people of varied ethnicity (for example, for economic reasons), leading to greater opportunity for racial frictions between ethnic groups and, consequently, increased racial bias reflected in the popularity of searching for racial slurs. Therefore, seeing the correlation between state-level racial slur search popularity and total population, state, Asian and Hispanic total population data is included in the dataset to investigate its relevance in predicting crime.

Table 3.5: Average Total Population and Average Monthly Search Popularity of Racial Slurs (State-level, 2004-2019)

U.S. State	Avg. Total Population	Rank*	Avg. Anti-Asian Slur Search Popularity	Rank**	Avg. Anti-Hispanic Slur Search Popularity	Rank*
California	37,701,500	1^{st}	11.23	2^{nd}	12.74	1^{st}
Texas	25,805,618	2^{nd}	8.64	4^{th}	10.12	2^{nd}
New York	19,416,148	3^{rd}	14.56	1^{st}	8.06	4^{th}
Florida	19,365,823	4^{th}	9.00	3^{rd}	8.71	3^{rd}
Illinois	12,768,541	5^{th}	8.09	5^{th}	7.52	7^{th}
Pennsylvania	12,683,688	6^{th}	7.08	9^{th}	7.92	6^{th}
Ohio	11,563,765	7^{th}	5.36	13^{th}	5.52	8^{th}
Michigan	9,957,641	8^{th}	6.27	11^{th}	5.17	9^{th}

*Out of a total of 51 due to the 50 U.S. state and 1 district (Washington D.C.)
**Out of a total of 49 due to the omittance of states with a lack of data, Wyoming and Vermont.
Note: Rankings are descending, the state with the highest value being ranked first.

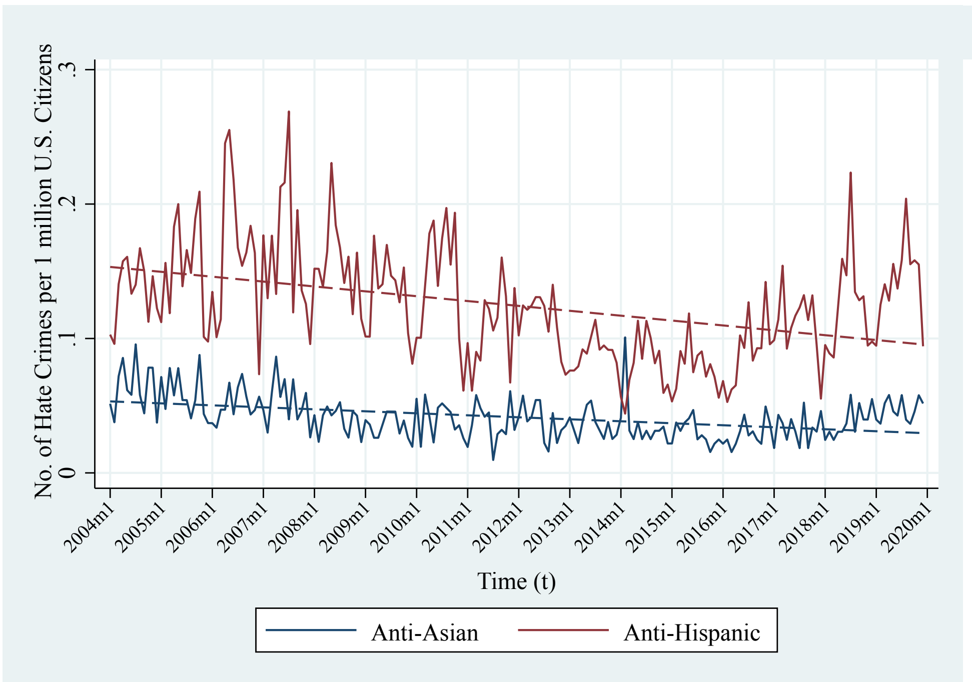
3.2 Hate Crime Data

As earlier mentioned in Section 2, defining "hate crime" has been widely discussed amongst scholars. The two largest datasets for hate crime data come from the FBI and BJS and will be discussed in Section 3.2.1 and 3.2.2, respectively.

3.2.1 Federal Bureau of Investigation

Data on reported hate crime from the FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.) database, spans the years 1995-2019 and includes victim ethnicity for each crime, which law enforcement agencies submit voluntarily. The data includes the location of committed hate crime (geographic location), type of location (e.g. school, college, home, convenience store), the reporting agency and its location, victim ethnicity and motivational bias of crime (e.g. sexual orientation, race, ethnicity, religion)[10]. Figure 3.5 graphs the total number of anti-Asian and -Hispanic hate crimes in the U.S. per one million people and their diminishing trend, based upon data from the FBI.

Figure 3.5: Number of Hate Crimes between 2004-2019 (National-level)

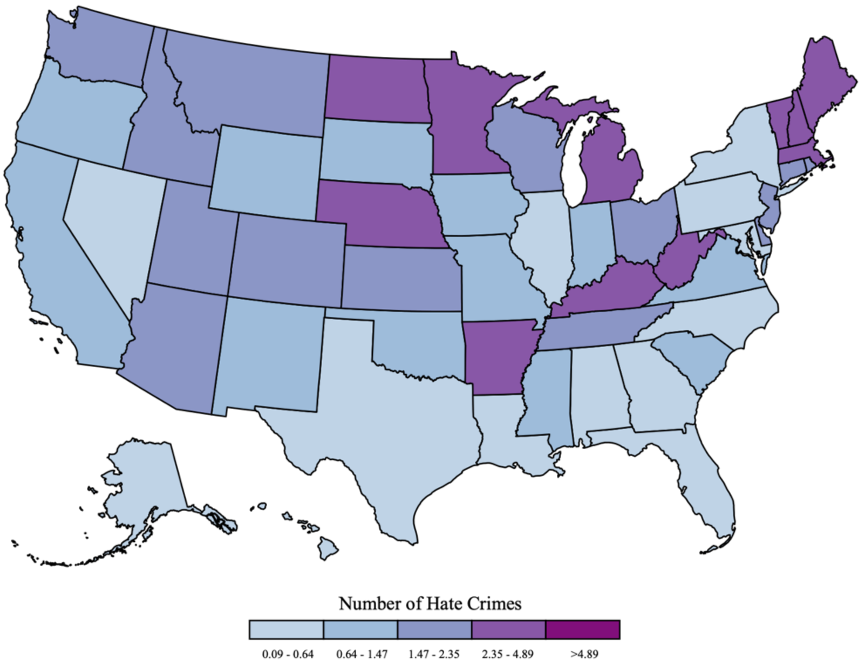


Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

[10] Appendix Tables A.1, A.2 and A.3 show an extensive list of crimes, motivational bias and location types.

The FBI, with respects to collecting hate crime data, considers "criminal offense against a person or property motivated in whole or in part by an offender's bias against a race, religion, disability, sexual orientation, ethnicity, gender, or gender identity" a hate crime (FBI, 2022). Similarly, in this research, crime which lists the offender as committing a crime with a motivational bias is considered a hate crime. However, in order to attribute the hate crime directly as being anti-Asian or anti-Hispanic, only crimes with a single motivational bias are considered i.e. crime which targets multiple ethnicities is excluded. Figures 3.6 and 3.7 show the average yearly number of hate crimes per 100,000 of Asian and Hispanic population over 2004-2019, respectively.

Figure 3.6: Average Yearly Number of Anti-Asian Hate Crime per 100,000 Asians (State-level, 2004-2019)

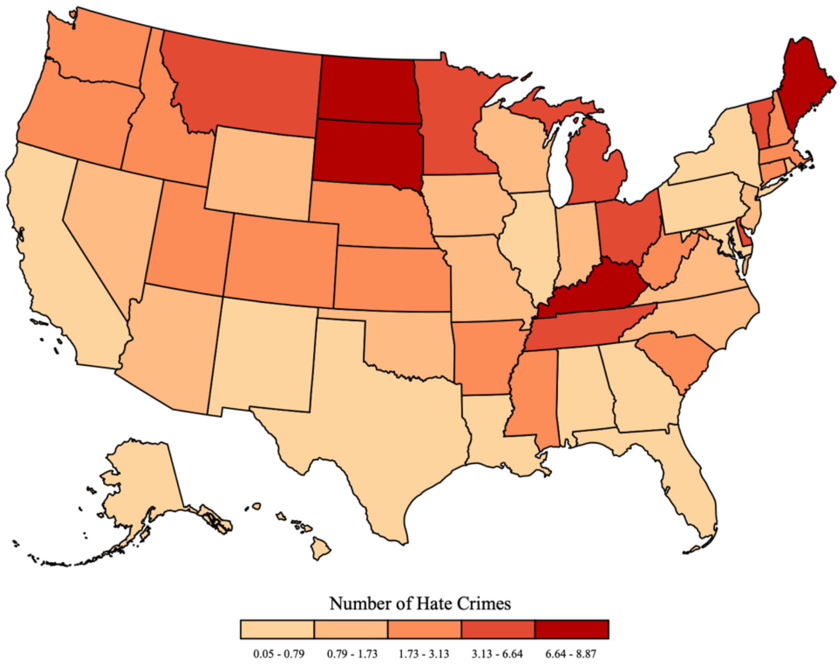


Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

North Dakota, Kentucky, Maine and Washington D.C. are among the states with the highest levels of relative hate crime against both ethnic minority groups. Again, without diving deeply into additional research, there seems to be no clear connection between a state's average yearly number of hate crimes and total population or ethnic population data. However, this is in-part controlled for by including population data for the main results. Table 3.6 and 3.7 list 8 states with the highest average yearly hate crime against Asians and Hispanics, respectively.

Future research might seek to identify reasons for why these states demonstrate above average levels of hate crimes against Asians and Hispanics (per 100,000 Asians or Hispanics).

Figure 3.7: Average Yearly Number of Anti-Asian Hate Crime for 100,000 Hispanics (State-level, 2004-2019)



Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

Table 3.6: Average Yearly Anti-Asian Hate Crime and Average Population (State-level, 2004-2019)

U.S. State	Avg. Yearly Anti-Asian Hate Crime per 100,000	Rank*	Avg. Asian Population (% of Total)	Rank*	Avg. Total Population	Rank*
Vermont	4.89	1st	1.43 %	40th	624,256	50th
Nebraska	4.16	2nd	2.04 %	30th	1,844,261	37th
Kentucky	4.01	3rd	1.26 %	43rd	4,347,627	26th
North Dakota	3.81	4th	1.17 %	45th	700,931	48th
Washington D.C.	3.75	5th	3.71 %	16th	630,209	49th
Massachusetts	3.69	6th	5.82 %	7th	6,637,334	14th
Maine	3.50	7th	1.07 %	47th	1,328,825	41st
New Hampshire	3.38	8th	2.37 %	28th	1,324,978	42nd

*Out of a total of 51 due to the 50 U.S. state and 1 district (Washington D.C.)
Note: Rankings are descending, the state with the highest value being ranked first.

Table 3.7: Average Yearly Anti-Hispanic Hate Crime and Average Population (State-level, 2004-2019)

U.S. State	Avg. Yearly Anti-Hispanic Hate Crime per 100,000	Rank*	Avg. Hispanic Population (% of Total)	Rank*	Avg. Total Population	Rank*
Washington D.C.	8.87	1st	9.84 %	19th	630,209	49th
South Dakota	8.65	2nd	3.23 %	43rd	827,476	46th
North Dakota	7.10	3rd	2.76 %	47th	700,931	48th
Kentucky	6.69	4th	3.05 %	46th	4,347,627	26th
Maine	6.68	5th	1.41 %	50th	1,328,825	41st
Vermont	5.62	6th	1.60 %	49th	624,256	50th
Montana	4.37	7th	3.28 %	42nd	1,001,357	44th
Tennessee	3.74	8th	4.57 %	37th	6,404,042	17th

*Out of a total of 51 due to the 50 U.S. state and 1 district (Washington D.C.)
Note: Rankings are descending, the state with the highest value being ranked first.

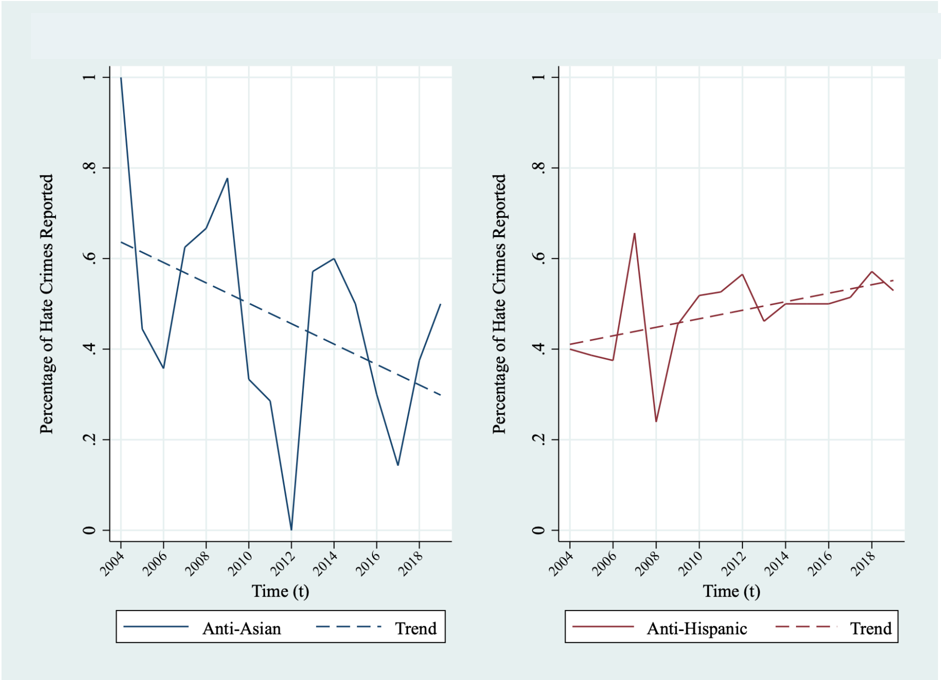
3.2.2 Bureau of Justice Statistics

A complementary source for hate crime data is the annual nationally representative National Crime Victimisation Survey (NCVS) conducted by the Bureau of Justice Statistics (BJS, n.d.), covering approximately 160,000 individuals over 2003-2019. Although this survey captures both reported and unreported hate crime, as well as other person- and household-specific data on education, income and demographic indicators, it fails to report in which city, county or state the hate crime is committed. Therefore, FBI data is preferrable for this research as it allows for the utilisation of geographical disaggregated racial bias data based on Google Trends.

The NCVS dataset, however, does deliver intriguing data about the difference in willingness to report hate crimes between Asians and Hispanics. In the survey, respondents are asked whether they have experienced a hate crime – a crime they suspect was committed against them due to their race, national origin, religion, disability, gender or sexual orientation.

Figure 3.8 highlights the disparity between Asian and Hispanic victims' willingness to report experienced hate crimes. Over 16 years of data and based upon data from the NCVS, Asians have become less likely to report hate crimes compared to Hispanics. However, the negative (positive) trend in reporting anti-Asian (anti-Hispanic) hate crime ought not to impact prediction error of the results in this current paper drastically as linear trends will be picked up by the training of the prediction model on sufficient in-sample data which includes these trends.

Figure 3.8: Willingness to Report Hate Crime (2004-2019)



Data Source: National Crime Victimisation Survey (NCVS) conducted by the Bureau of Justice Statistics (BJS, n.d.)

3.3 Population & Labour Data

As research indicates that hate crime levels are likely to depend on the proportion of a region's population which consists of the victimised ethnic group (Green et al., 1998; Espiritu, 2004; King et al., 2007; Piatkowska et al., 2018; Wagner et al., 2020) population data is included. The United States Census Bureau (US Census Bureau, 2022) offers annual total population estimate data on town and city, county, state and national level dating back to 1980. This includes county and state-level population data over the Asian and Hispanic population.

Monthly population data points (x^m) are estimated from annual population data points (x^a) using linear interpolation (Equation 3.2) and extrapolation (Equation 3.3).

$$x_t^m = x_{t_0}^a + (t - t_0)\frac{(x_{t_1}^a - x_{t_0}^a)}{(t_1 - t_0)}, \qquad t = \{t_0, \dots, t_1\} \tag{3.2}$$

Where $x_{t_0}^a$ and $x_{t_1}^a$ are two annual data points and t is measured in months. If the two nearest data points to $x_{t^*}^m$ are $x_{t_0}^a$ and $x_{t_1}^a$ then extrapolation can be formulated as:

$$x_{t^*}^m = x_{t_0}^a + (t^* - t_0)\frac{(x_{t_1}^a - x_{t_0}^a)}{(t_1 - t_0)} \tag{3.3}$$

One caveat of implementing linear inter- and extrapolation is that new imputed data can contain estimation error if the original data suffers from seasonal or cyclical trends. The estimation error is believed to be negligible as the U.S. population has grown steadily across 2004-2019.

Monthly labour data on employment between 2004 and 2019 is compiled by the U.S. Department of Agriculture (USDA, 2021), using data straight from the Bureau of Labour Statistics (BLS) Local Area Unemployment Statistics (LAUS). Labour force participation, total employment and total unemployment data is combined with population data to construct rates.

3.4 Data Transformation

Section 3.4.1 discusses the need for the standardisation of data before using LASSO. To improve prediction accuracy of the prediction models, data is tested for non-stationary and transformed, the methodology for this is presented in Section 3.4.2.

3.4.1 Standardisation

As LASSO penalises predictors "unfairly" dependent on the magnitude of predictors' coefficients in the regression, predictors need to be standardised. If predictors are not standardised, predictors with large absolute values such as population data will have smaller coefficients which will be penalised less through the LASSO method and are more likely to remain from the selected prediction model. Conversely, predictors with small absolute values such as Google Trends data will have larger coefficients which will be penalised more through the LASSO method and are more likely to be removed from the selected prediction model – this becomes clearer in Section 4.1.

Equation 3.4 shows a general standardisation of a predictor:

$$z = \frac{x - \mu}{\sigma} \tag{3.4}$$

where x is the variable to be standardised, μ its mean and σ its standard deviation. The general standardisation is adapted in an attempt to avoid contamination between in- and out-of-samples and implemented for both national- and state-level data. To do this, predictors are standardised using only the in-sample mean (μ_{IN}) and standard deviation (σ_{IN}), rather than overall observations. This avoids out-of-sample observations containing in-sample information and biasing out-of-sample performance. Equation 3.5 shows the time-series standardisation, whilst Equation 3.6 shows the state-level data standardisation:

$$z_t = \frac{x_t - \mu_{IN}}{\sigma_{IN}}, \qquad t = \{1, \dots, T\} \tag{3.5}$$

$$z_{i,t} = \frac{x_{i,t} - \mu_{i,IN}}{\sigma_{i,IN}}, \qquad t = \{1, \dots, T\}; \qquad i = \{1, \dots, n\} \tag{3.6}$$

3.4.2 Stationarity

Van Greulen et al. (2014, p. 1) and Gujarati (2009, pp. 380-382) highlight the necessity of using stationary data for accurate forecasting. Hendry & Pretis (2016) emphasize that failing to account for non-stationary data will lead to poor time-series prediction accuracy.

A stationary time series is one who's statistical properties, such as mean and variance, do not change over time i.e. one that is void of seasonal or cyclical trends and is predictable in the long-term. Predicting stationary data is much easier than non-stationary data as its statistical properties do not change over time – reduced external variability.

A data series can be tested for stationarity using a unit root test. A unit root exists if the variance of a stochastic process (e.g. a random walk with a drift) is dependent on time. The existence of a unit root concludes that data is non-stationary.

Using a time-series random walk:

$$y_t = a_1 y_{t-1} + \varepsilon_t \tag{3.7}$$

Where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ and $t = \{1,2,3, \dots, T\}$. Assuming $a_1 = 1$:

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t \\ &= y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &\ \ \vdots \\ y_t &= y_0 + \sum_{i=1}^{t} \varepsilon_i \end{aligned}$$

Taking the variance of y_t leads to:

$$Var(y_t) = Var\left(\sum_{i=1}^{t} \varepsilon_i\right) = t\sigma^2 \tag{3.8}$$

As Equation 3.8 shows that the stochastic process' variance is time-dependent, we can conclude that a unit root exists and thus we have non-stationary data. Therefore, to make the variance of the data used in this paper independent of time, it is necessary to transform its data.

The modified Dickey-Fuller t-test (DF-GLS) by Elliott et al. (1996) is used to test for unit roots in the time-series data. The DF-GLS test's null hypothesis is that y_t is non-stationary (random walk or random walk with a drift) with the alternative hypothesis that y_t is stationary around a linear trend. Before performing the test, the modified version of the original Dickey-Fuller (DF) t-test transforms time-series data using a generalised least squares (GLS) regression. The DF-GLS t-test is used because academics, in addition to the original authors of the DF-GLS t-test, such as Vougas (2008) and Westerlund (2013) show the DF-GLS to be more effective in identifying unit roots than DF, especially with a small sample such as the time-series data in this research. Testing for unit roots in the time-series data using the DF-GLS t-test is done as to identify which variables need data transformation. The unit root test is conducted again after data transformation to assess whether the variable's stationarity.

Data transformation can be used to help increase data stationarity. For this research, first- and second-differencing, logging and the combination of logging and then first-differencing is used to improve prediction results. If data does not produce a more stationary variable after transformation, the non-transformed variable is used for prediction selection to allow for the selection of the most stationary variables possible. Transformed variables can make the interpretation of results for causality slightly more difficult, often needing a back transformation (Lee, 2020). However, as this research is focused less on inference and more on the power to predict change in hate crime (the dependent variable) rather than a parametric analysis, a backward transformation will not be needed to interpret results.

Table 3.8 shows unit root test results for national-level data before and after data transformation. Transforming the data significantly increases the number of stationary variables from 27 to 57. It must be noted that the transformation of the dependent variables, anti-Asian and -Hispanic hate crime, fails to reject the presence of non-stationarity (null hypothesis). The number of non-stationary dependent and independent variables make prediction of anti-Asian and -Hispanic hate crime more difficult and, thus, lead to more inaccurate results which is discussed Section 7.1.

Table 3.8: t-Test Statistic from the DF-GLS Stationarity Test (National-level)

Variables	Before Transformation	After Transformation	Variables	Before Transformation	After Transformation
chink	-3.8575**	-3.8575**	*coolie*^{2}	-3.6588**	-3.6588**
chink^2	-3.4896**	-3.4896**	*coolies*	-2.8614*	-2.8614*
chinks	-2.0358	-2.8047*	*coolies*^{2}	-4.3119**	-4.3119**
chinks^2	-2.6130	-1.6548	*chinky*	-1.6466	-2.7578
gook	-3.0050*	-3.0050*	*chinky*^{2}	-1.8954	-2.4856
gook^2	-3.4919**	-3.4919**	*chinkys*	-2.3852	-4.2674**
gooks	-1.6575	-5.1222**	*chinkys*^{2}	-1.9459	-4.5673**
gooks^2	-2.2329	-5.2760**	*wetback*	-2.6669	-2.8595*
hapa	-2.2961	-2.6939	*wetback*^{2}	-1.6133	-3.3474*
hapa^2	-2.3725	-2.8933*	*wetbacks*	-2.3295	-4.2616**
hapas	-1.1878	-1.3667	*wetbacks*^{2}	-3.2724*	-3.2724*
hapas^2	-1.3514	-1.4127	*beaner*	-2.9136*	-2.9136*
jap	-2.4716	-3.1865*	*beaner*^{2}	-2.9213*	-2.9213*
jap^2	-2.4310	-3.5468**	*beaners*	-1.1608	-2.8504*
japs	-0.8442	-3.6660**	*beaners*^{2}	-1.4521	-2.0864
japs^2	-0.7586	-6.2118**	*chopa*	-2.5609	-5.3295**
paki	-1.9013	-3.0226*	*chopa*^{2}	-3.6974**	-3.6974**
paki^2	-1.9144	-2.9916*	*chopas*	-2.0717	1.0279
pakis	-1.8727	-1.4796	*chopas*^{2}	-2.4100	-4.3164**
pakis^2	-3.3706*	-1.0047	*chopita*	-2.3422	-5.5390**
chinaman	-1.8425	-2.1187	*chopita*^{2}	-2.8263*	-2.8263*
chinaman^2	-1.3901	-4.4878**	*pepper belly*	-2.2493	-1.8894
chinamen	-3.1475*	-3.1475*	*pepper belly*^{2}	-2.8197*	-2.8197*
chinamen^2	-3.8968**	-3.8968**	*pepper bellys*	-3.3639*	-3.3639*
ching chong	-1.6691	-0.1651	*pepper bellys*^{2}	-4.6694**	-4.6694**
ching chong^2	-2.2483	-2.4183	*taco bender*	-3.1281*	-3.1281*
ching chongs	-10.3601**	-10.3601**	*taco bender*^{2}	-3.4959**	-3.4959**
ching chongs^2	-57.8180**	-57.8180**	*spic*	-1.2075	-0.6150
dragon lady	-1.5748	-2.6960	*spic*^{2}	-1.1213	-1.4908
dragon lady^2	-2.0098	-2.0623	*spics*	-2.2541	-4.1792**
dragon ladies	-1.9431	-1.7637	*spics*^{2}	-2.1703	-4.5732**
dragon ladies^2	-2.8254*	-2.8254*	*Total Employment*	-1.9948	-2.2211
buddha head	-2.2802	-1.6969	*Employment (%)*	-1.7559	-2.3176
buddha head^2	-2.8817*	-2.8817*	*Total Labour Force Participation*	-1.5665	-2.2700
buddha heads	-2.0233	-4.9711**	*Labour Force Participation (%)*	-1.4125	-3.1462*
buddha heads^2	-2.0492	-6.013**	*Total Unemployment*	-1.7233	-2.2423
squint eye	-3.4065*	-3.4065*	*Unemployment (%)*	-1.6411	-3.9897**
squint eye^2	-3.6293**	-3.6293**	*Total Asian Population*	-1.9006	-2.2778
squint eyes	-2.1219	-0.2760	*Asian Population (%)*	-0.7494	-3.3072*
squint eyes^2	-2.6820	-1.8768	*Total Hispanic Population*	-2.1067	-2.5252
chinee	-1.8456	-4.2865**	*Hispanic Population (%)*	-1.5727	-4.3719**
chinee^2	-1.8041	-4.0661**	*Total Population*	-0.9978	-2.3672
chinees	-2.0524	-2.2423	*Anti-Asian Hate Crime*	-1.7686	-2.1589
chinees^2	-1.8040	-5.0147**	*Anti-Hispanic Hate Crime*	-1.0545	-1.4882
coolie	-3.1388*	-3.1388*			

No. of Stationary Variables Before Data Transformation 27

No. of Stationary Variables After Data Transformation 57

** - test-statistic exceeds 1% critical level.

* - test-statistic exceeds 5% critical level.

Stationarity in the state-level data is tested using the Levin-Lin-Chu Test (Levin et al., 2002) as it fares well with "moderate" sized panel data – between 10 to 250 individuals and 25 to 250 temporal observations per individual. The Levin-Lin-Chu test's null hypothesis of a unit root, similar to the DF-GLS test, assumes non-stationarity in y_t, whilst the alternative hypothesis conjectures y_t to be stationary. The state-level unit root test results for before and after data transformation are shown by Table 3.9. After data transformation, all variables significantly reject the presence of a unit root i.e. all variables are stationary – the necessary condition for meaningful prediction models.

Table 3.9: p-values from the Levin-Lin-Chu Stationarity Test (State-level)

Variables	Before Transformation	After Transformation	Variables	Before Transformation	After Transformation
gook	0.0000**	0.0000**	*beaners*^{2}	0.0000**	0.0000**
gook^2	0.0000**	0.0000**	*spic*	0.0000**	0.0000**
gooks	0.0000**	0.0000**	*spic*^{2}	0.0000**	0.0000**
gooks^2	0.0000**	0.0000**	*spics*	0.0000**	0.0000**
ching chong	0.0000**	0.0000**	*spics*^{2}	0.0000**	0.0000**
ching chong^2	0.0000**	0.0000**	*Total Employment*	0.0000**	0.0000**
japs	0.0000**	0.0000**	*Employment (%)*	0.3042	0.0000**
japs^2	0.0000**	0.0000**	*Total Labour Force Participation*	0.0001**	0.0001**
paki	0.0000**	0.0000**	*Labour Force Participation (%)*	0.0064**	0.0000**
paki^2	0.0000**	0.0000**	*Total Unemployment*	0.0000**	0.0000**
pakis	0.0000**	0.0000**	*Unemployment (%)*	0.9888	0.0000**
pakis^2	0.0000**	0.0000**	*Total Asian Population*	0.9621	0.0000**
wetback	0.0000**	0.0000**	*Asian Population (%)*	0.6930	0.0000**
wetback^2	0.0000**	0.0000**	*Total Hispanic Population*	0.0000**	0.0000**
wetbacks	0.0000**	0.0000**	*Hispanic Population (%)*	0.0000**	0.0000**
wetbacks^2	0.0000**	0.0000**	*Total Population*	0.0000**	0.0000**
beaner	0.0000**	0.0000**	*Anti-Asian Hate Crime*	0.0000**	0.0000**
beaner^2	0.0000**	0.0000**	*Anti-Hispanic Hate Crime*	0.0000**	0.0000**
beaners	0.0000**	0.0000**			

No. of Stationary Variables Before Data Transformation 33

No. of Stationary Variables After Data Transformation 37

** - test-statistic exceeds 1% critical level.

* - test-statistic exceeds 5% critical level.

4 METHODOLOGY

Section 4.1 and 4.2 introduce the applied methodology in the general-setting and how it is implemented in the statistical package utilised in this paper. Section 4.3 extends on the previous two sections by introducing the application of the methodology in a time-series (national-level) and panel data (state-level) setting and proposes prediction models resulting from the solving of minimisation problems.

4.1 General LASSO

Regression models do not just help with identifying causal variable relationships, but also offer a basis for prediction. Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996) is a method used for selecting and fitting variables in order to compute a prediction model by providing a disciplined mechanism for a bias-variance trade-off i.e. balancing between over- and underfitting of the data. Research from Nitta et al. (2019), Wang et al. (2020) and Fatehkia et al. (2019) have implemented LASSO in order to help with accuracy and variable selection in crime prediction. With LASSO, we choose $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, the coefficient parameter vector, which minimises $(y_i - \beta_0 - \boldsymbol{x}'_i\boldsymbol{\beta})^2$, where β_0 is the constant term, y_i denotes the outcome variable and $\boldsymbol{x}'_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ a vector of all independent variables as subject to $\sum_{j=1}^{p}|\beta_j| \leq t$ – with p equalling the total number of predictors and the tuning parameter t controlling the magnitude of shrinkage that is applied on regression coefficient estimates. Equations 4.1 and 4.2, based on work by Tibshirani (1996), express the above mentioned minimisation:

$$\mathbb{E}[(y_i - \beta_0 - \boldsymbol{x}'_i\boldsymbol{\beta})^2] \qquad s.t. \quad \sum_{j=1}^{p}|\beta_j| \leq t \tag{4.1}$$

Rewriting the above, heralds the following:

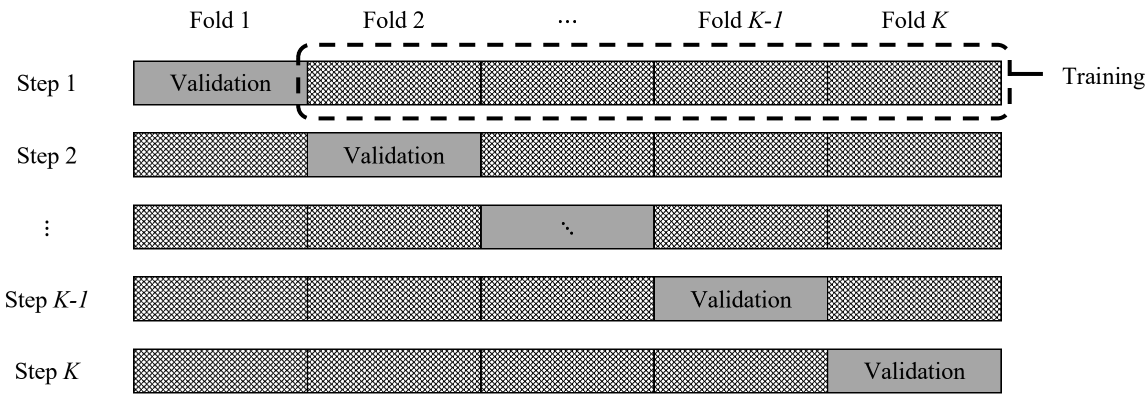
$$\min_{\beta_0, \beta}\left\{\sum_{i=1}^{N}[(y_i - \beta_0 - \boldsymbol{x}'_i\boldsymbol{\beta})^2]\right\} + \lambda\sum_{j=1}^{p}|\beta_j| \tag{4.2}$$

where λ is effectively the penalisation cost for selecting a coefficient parameter (β) not equal to zero. Since the optimal penalisation parameter (λ) which minimises the mean squared error (MSE) – sometimes called the "LASSO prediction error" – is unknown, cross-validation (CV) is used to select the optimal penalisation parameter. Before CV is implemented the dataset is divided into "training set" (in-sample) and "test set" (out-of-sample) – the exact division is further defined in Section 4.3. The training set is further split into so-called "training sample" and "validation sample" in order for CV to select the optimal penalisation parameter λ from which to build a prediction model and calculate an in-sample MSE. Finally, the same prediction model is evaluated based on its out-of-sample prediction performance (MSE) using the test set.

4.2 lassopack: General Setting

Prediction results are attained by using the statistical package *lassopack*, created by Ahrens et al. (2020) for STATA. In a general setting, the command *cvlasso* implements a K-fold CV LASSO (Figure 4.1) which splits the training set into K number of almost equally sized samples, also denoted as "folds", with n_k expressing the number of folds (where $k = 1, \dots, K$).

Figure 4.1: Graphical depiction of K-fold CV



Note: Adapted from "lassopack: Model selection and prediction with regularized regression in Stata" by A. Ahrens, C. Hansen, M. Schaffer, 2020, p. 187. (https://doi.org/10.1177/1536867x20909697).

In a K-fold CV, Fold 1 ($k = 1$) would be assigned to be the validation sample in Step 1, with the other $K - 1$ number of folds constructing the training sample. With each new step, the next fold is assigned to be the validation sample, until each fold, therewith every set of observations (Ω_k), has been used for validation at least once. Therefore, the number of steps must equal the number of folds. In each step, and for given values of the penalisation (λ), the model is fitted to the training sample, subsequently an estimate is produced for each step, denoted as $\hat{\beta}_k(\lambda)$. For a single fold the mean square prediction error (MSPE) is constructed as

$$MSPE_k(\lambda) = \frac{1}{n_k} \sum_{i \in \Omega_k} \left\{ y_i - \boldsymbol{x'}_i \widehat{\boldsymbol{\beta}}_k(\lambda) \right\}^2 \tag{4.3}$$

Finally, the prediction performance, assuming that training and validation samples are serially independent, is measured by averaging MSPE across all folds, resulting in

$$\hat{\mathcal{L}}^{CV}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} MSPE_k(\lambda) \tag{4.4}$$

Even when performing LASSO, keeping α fixed, computing $\hat{\mathcal{L}}^{CV}$ for different values of λ till the prediction performance is maximised can be computationally costly. To minimise the prediction error a 10-fold CV is used, as research from Hastie et al. (2009) and Arlot and Celisse (2010) argues that using a K value above 10 seldomly improves predictive power.

4.3 lassopack: Time-series and Panel Data Setting

As *cvlasso* is used on time-series (national-level) and panel (state-level) data, it is important to discuss how the STATA command accommodates this change in setting. The implemented method, referred to as "rolling h-step-ahead cross-validation", assigns training and validation samples from the "training set" (in-sample) within the time-series framework, best explained through Figure 4.2[11]. Each set has a fixed number of samples with the last sample in the set being assigned for validation and all previous samples for training. With each new step the entire set moves up one time period.

For the main results, the years 2004 to 2016 are assigned to training and validating the models using *cvlasso* (in-sample), whilst 2017 to 2019 are used to test the out-of-sample prediction performance. The more data the training and validation process receives the higher the chance it produces accurate models, therefore, it has been given 81.25% of the dataset, whilst the remaining 18.75% is used to assess the models' out-of-sample performance.

The largest challenge faced by implementing this method is the violation of independency between training and validation samples (Ahrens et al., 2020). As my research uses dependent data its results will be optimistically downward biased, showing a lower prediction error compared to the true error. However, research from Bergmeir et al. (2018) shows that K-fold CV remains applicable on autoregressive models in the time-series setting as long as the assumption is made that errors are uncorrelated. The existing bias of the model is addressed as part of the robustness checks in Section 6.1. In short, however, *cvlasso* is applied on an autoregressive process of order 6 to produce an unbiased benchmark performance. This benchmark model is compared with the biased main model to examine whether Google Trends data adds to predicting hate crime more accurately, keeping in mind the existing bias.

---

[11] The "h-step-ahead CV" approach has similarities to research from Burman, Chow & Nolan (1994), who introduced "h-block CV" as an extension of "Leave-One-Out CV", but implemented with classic CV.

Figure 4.2: Rolling *h*-step-ahead Cross Validation (Fixed Window *h* = 1)

		Step				
		1	2	3	4	5
	1	T	·	·	·	·
	2	T	T	·	·	·
	3	V	T	T	·	·
t	4	·	V	T	T	·
	5	·	·	V	T	T
	6	·	·	·	V	T
	7	·	·	·	·	V

Note: Adapted from "lassopack: Model selection and prediction with regularized regression in Stata" by A. Ahrens, C. Hansen, M. Schaffer, 2020, p. 189. (https://doi.org/10.1177/1536867x20909697).

Turning to the setup for the main results, the national-level (time-series) problem, for a given forecasting period f, to be solved by LASSO is:

$$\min_{\beta_0,\beta}\left\{\sum_{t=0}^{T}\left[\left(\Delta y_{t+f}-\beta_0-\boldsymbol{x}'_t\boldsymbol{\beta}\right)^2\right]\right\}+\lambda\sum_{j=1}^{p}\left|\beta_j\right|, \qquad f=(0,1,\ldots,6),\ t=(0,1,\ldots,T) \tag{4.5}$$

Where $\Delta y_{t+f}=(y_{t+f}-y_{(t-1)+f})$ is the absolute change in the dependent variable, anti-Asian or -Hispanic hate crime, in time period $t+f$. This means that 7 different prediction models are made, one for each forecasting period. The independent variable vector $\boldsymbol{x}'_{,t}$ includes racial slur search popularity, population and labour variables.

The state-level (panel data) problem is set up similarly to that of the national-level (time-series) problem (Equation 4.5) with the only addition being that it is done per state, s:

$$\min_{\beta_0,\beta}\left\{\sum_{s=1}^{S}\sum_{t=0}^{T}\left[\left(\Delta y_{s,t+f}-\beta_0-\boldsymbol{x}'_{s,t}\boldsymbol{\beta}\right)^2\right]\right\}+\lambda\sum_{j=1}^{p}\left|\beta_j\right|, \qquad s=(1,\ldots,S) \tag{4.6}$$

The solving of the time-series and panel data minimisation problems will lead to prediction models such as Equation 4.7 and 4.8, respectively.

$$\Delta y_{t+f}=\widehat{\beta_0}+\dot{\boldsymbol{x}}'_t\widehat{\boldsymbol{\beta}}+\varepsilon_t \tag{4.7}$$

$$\Delta y_{s,t+f} = \widehat{\beta_0} + \dot{\boldsymbol{x}}'_{s,t}\widehat{\boldsymbol{\beta}} + \varepsilon_{s,t} \tag{4.8}$$

Here, $\dot{\boldsymbol{x}}'$ is a vector of all the independent variables selected by LASSO. Respectively, $\widehat{\beta_0}$ and $\widehat{\boldsymbol{\beta}}$ are the LASSO estimated constant term and estimated coefficient parameter vector with ε being the error term. Finally, the LASSO prediction models are compared to simple benchmark models consisting of autoregressive processes of order 6 (AR6) and moving average processes of order 3 (MA3).

5 RESULTS

National-level results are presented in Section 5.1, whilst in Section 5.2 the state-level results are discussed. Ultimately, Section 5.3 focuses on the comparison of out-of-sample performance of LASSO constructed prediction models, from the previous sections, with benchmark models. The results are further analysed and discussed in Section 7.1.

In ensuing sections, models show how well today's monthly data (f = 0) predicts anti-Asian and -Hispanic hate crime on the national-level and state-level for future periods:

Model (1)	today (f = 0);	Model (5)	in 4 month (f = 4);
Model (2)	in 1 month (f = 1);	Model (6)	in 5 month (f = 5);
Model (3)	in 2 month (f = 2);	Model (7)	in 6 month (f = 6);
Model (4)	in 3 month (f = 3);		

5.1 National-level Results

Generally, we expect in-sample prediction error, assessed by the root-square-mean error (RMSE), to be lower than out-of-sample. Turning to the national-level prediction results for anti-Asian hate crime (Table 5.1), however, this is only seen for the prediction models (1), (5) and (6). These three models compared to the other four models, which better predict out-of-sample than in-sample, in respect to the number of selected predictors, include a significantly higher number of predictors. Model (1) explains 46.17% of the in-sample variation in the dependent variable, whereas the other models (2-7) capture much less variation – between 0 to 4.8%. With the selection of every possible predictor, very good in-sample, but poor out-of-sample prediction performance, Model (1) shows signs of overfitting – the very low optimal penalisation coefficient (λ) of 0.03, compared to that chosen in the other models, adds to the possibility of overfitting. Some racial slurs are selected more often than other slurs such as *gooks*, *ching chongs*, *coolie* and *chinky(s)* are included in three models.

Table 5.1: Anti-Asian Results (National-level, 2004-2019)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
chink	0.1419						-0.0111
chink^2	-0.0030					-0.0791	
chinks	0.1193						-0.1262
chinks^2	0.1179						
gook	0.1657						
gook^2	-0.2688						
gooks	0.2585					-0.1137	0.0464
gooks^2	-0.1643						
hapa	0.1527						
hapa^2	-0.4172						
hapas	-0.0736						-0.0096
hapas^2	0.0748						
jap	1.1922						
jap^2	-1.5064						
japs	-0.0222						
japs^2	0.1065						0.0281
paki	0.2132						
paki^2	-0.5507						0.0463
pakis	0.0671						0.0294
pakis^2	-0.2013						
chinaman	-0.0429					0.0089	
chinaman^2	0.1619						
chinamen	0.0791						
chinamen^2	-0.0920						
ching chong	-0.0058					-0.0285	0.0532
ching chong^2	0.0438						
ching chongs	-0.1090						
ching chongs^2	0.3575						0.0423
dragon lady	-0.1658						0.0883
dragon lady^2	0.0957						
dragon ladies	-0.1736						-0.0555
dragon ladies^2	-0.0956						
buddha head	-0.3782						
buddha head^2	0.1226						
buddha heads	-0.0053						0.0319
buddha heads^2	0.1620						
squint eye	0.1752						
squint eye^2	0.0067						-0.0600
squint eyes	-0.2126						
squint eyes^2	0.3465						
chinee	0.1872					-0.0914	0.1178
chinee^2	-0.1628						
chinees	0.2515						
chinees^2	-0.3486						0.0644
coolie	-0.6819					-0.0246	0.0065

coolie²	0.8615						
coolies	0.2785						
coolies²	-0.2349					0.0781	-0.0112
chinky	-0.1863					0.0062	-0.0308
chinky²	0.2290						
chinkys	0.4159					0.1155	-0.0975
chinkys²	-0.5086	0.0401					
Total Employment	0.2521						
Employment (%)	0.3397						
Total Labour Force Participation	-0.1316						0.0086
Labour Force Participation (%)	-0.4074	0.0527				-0.0047	0.0171
Total Unemployment	0.0002						
Unemployment (%)	0.1698					-0.0147	0.0560
Total Asian Population	0.6754						
Asian Population (%)	-0.5559						
Total Hispanic Population	-4.0333						
Hispanic Population (%)	3.0232						
Total Population	0.4284						
Anti-Asian Hate Crime (Lag 1)	-0.6953					-0.0220	0.0197
Anti-Asian Hate Crime (Lag 2)	-0.5849						
Anti-Asian Hate Crime (Lag 3)	-0.4372						
Anti-Asian Hate Crime (Lag 4)	-0.4715						-0.0434
Anti-Asian Hate Crime (Lag 5)	-0.3278						
Anti-Asian Hate Crime (Lag 6)	-0.3544						0.1047
Constant	-0.0593	-0.0094	0.0045	-0.0090	-0.0141	0.0069	-0.0175
Selected Predictors	69	2	0	0	0	12	25
Selected Optimal λ	0.03	48.36	50.80	44.16	64.70	21.26	11.85
in-sample R^2	0.7163	0.0360	0.0000	0.0413	0.0000	0.1316	0.1881
in-sample Adj. R^2	0.4617	0.0161	-0.0068	0.0281	-0.0068	0.0480	0.0150
in-sample RMSE	0.5252	0.9597	0.9824	0.9522	0.9740	0.8963	0.8580
out-of-sample RMSE	**1.0879**	**0.8745**	**0.8485**	**0.8713**	**0.8649**	**0.9053**	**1.0024**
out-of-sample AR6 RMSE	**0.6468**	**0.9119**	**0.8767**	**0.8865**	**0.8910**	**0.8915**	**0.9027**
out-of-sample MA3 RMSE	**0.7704**	**0.8806**	**0.8533**	**0.8601**	**0.8646**	**0.8668**	**0.8613**

Turning to anti-Hispanic national-level results in Table 5.2, we see similarities to the anti-Asian results. Models (1), (3), and (5), models with more selected variables than the other models, have a lower in-sample than out-of-sample RMSE and much lower chosen optimal λ – this, too, may hint at overfitting on training data. Although, the remaining models (2), (4), (6) and (7) have no more than two selected predictors which may be the consequence of the extremely high selected optimal penalization coefficient (λ), they have the lowest out-of-sample prediction errors. Here, model (1) captures about 20.75% less in-sample variation in the dependent variable (25.42%) compared to the anti-Asian model (1). Population and employment predictors are not selected for any anti-Hispanic hate crime prediction model. All anti-Hispanic racial slurs are selected at least for two models.

Table 5.2: Anti-Hispanic Results (National-level, 2004-2019)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
wetback	-0.0108		0.0112		-0.0013		
wetback²	0.0002		-0.0002				
wetbacks			-0.0046		-0.0016		
wetbacks²	0.0001				-0.0001		
beaner	0.0220				0.0043		
beaner²	-0.0002		-0.0001		0.0001		
beaners	0.0003		0.0107		0.0141		
beaners²	0.0001		-0.0001	-0.0001	-0.0001		
chopa	-0.0026		-0.0088		-0.0024		
chopa²			0.0001		0.0001		
chopas	-0.0039		-0.0023		0.0007		
chopas²	0.0001				-0.0001		
chopita	-0.0067		0.0009		0.0131		
chopita²			-0.0001				
pepper belly	0.0179		0.0016		-0.0006		
pepper belly²	-0.0004		0.0002		0.0001		
pepper bellys	-0.0150		-0.0245		-0.0087		
pepper bellys²	0.0001		0.0002	-0.0001	0.0000		
taco bender	0.0048		0.0066		0.0210		
taco bender²			-0.0001		-0.0002		
spic			-0.0005		0.0004		
spic²							
spics	0.0103		0.0186		0.0152		
spics²	-0.0001		-0.0002		-0.0001		
Total Employment							
Employment (%)							
Total Labour Force Participation							
Labour Force Participation (%)							
Total Unemployment							
Unemployment (%)							
Total Asian Population							
Asian Population (%)							
Total Hispanic Population							
Hispanic Population (%)							
Total Population							
Anti-Hispanic Hate Crime (Lag 1)	-0.2496		-0.0330		-0.1584		
Anti-Hispanic Hate Crime (Lag 2)			0.0270		-0.0465		
Anti-Hispanic Hate Crime (Lag 3)	-0.0203		-0.1752		-0.1821		
Anti-Hispanic Hate Crime (Lag 4)			-0.0525				
Anti-Hispanic Hate Crime (Lag 5)	-0.0629		-0.1490				
Anti-Hispanic Hate Crime (Lag 6)							
Constant	-0.5971	-0.0104	0.0943	0.0202	-0.3411	-0.0011	0.0023
Selected Predictors	20	0	24	2	23	0	0
Selected Optimal λ	18.93	146,923.20	11.67	44,871.12	10.61	202,138.00	378,034.30
in-sample R^2	0.3903	0.0000	0.2315	0.0706	0.2369	0.0000	0.0000
in-sample Adj. R^2	0.2542	-0.0068	0.0442	0.0381	0.0588	-0.0068	-0.0068
in-sample RMSE	0.7887	1.0084	0.8841	0.9733	0.8885	1.0165	1.0165
out-of-sample RMSE	**1.0032**	**0.9635**	**1.2221**	**0.9908**	**1.1950**	**0.9474**	**0.9652**
out-of-sample AR6 RMSE	**0.9116**	**0.9703**	**0.9797**	**0.9619**	**0.9421**	**0.9482**	**0.9848**
out-of-sample MA3 RMSE	**0.9751**	**0.9647**	**0.9718**	**0.9683**	**0.9159**	**0.9450**	**0.9536**

5.2 State-level Results

As mentioned in Section 3 (Table 3.1), the state-level main results are based upon data from 35 U.S. states and 1 Federal District. Model (1) from Table 5.3 is able to explain 42.51% of the variation in the change of anti-Asian hate crime in $t + f = 0$. All other models have a negative adjusted R^2 exhibiting their failure to explain any in-sample variation of the anti-Asian hate crime. While there are similarities to the national-level model (1) (Table 5.1), such as the high number of selected predictors, the better out-of-sample than in-sample prediction performance of model (1) suggests the absence of overfitting of the model on the training sample. Models (2) through (5) have no selected predictors and higher optimal penalisation coefficients selected through CV than the other prediction models, which may be the cause for the low selection of predictors. Apart from *ching chong,* all racial slurs, including squared counterparts, are selected in at least one model.

Table 5.3: Anti-Asian Results (State-level, 2004-2019)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
gook						0.0083	-0.0070
gook²	-0.0029						
gooks	-0.0001						
gooks²						-0.0091	
ching chong							
ching chong²	-0.0184						
japs	0.0137						
japs²	0.0067						
paki	-0.0076					-0.0140	
paki²							0.0061
pakis						0.0012	
pakis²	-0.0067					0.0011	
Total Employment							
Employment (%)							
Total Labour Force Participation							
Labour Force Participation (%)	0.0023						0.0042
Total Unemployment							
Unemployment (%)	0.0020						
Total Asian Population							
Asian Population (%)	0.0034						
Total Hispanic Population	-0.0135						
Hispanic Population (%)							
Total Population							
Anti-Asian Hate Crime (Lag 1)	-0.8330						-0.0110
Anti-Asian Hate Crime (Lag 2)	-0.6778					-0.0077	
Anti-Asian Hate Crime (Lag 3)	-0.5281						0.0070
Anti-Asian Hate Crime (Lag 4)	-0.3574					0.0148	-0.0121
Anti-Asian Hate Crime (Lag 5)	-0.2352						
Anti-Asian Hate Crime (Lag 6)	-0.1230						0.0036
Constant	-0.0041	-0.0021	0.0005	-0.0021	-0.0028	0.0005	-0.0013

Selected Predictors	17	0	0	0	0	7	7
Selected Optimal λ	32.59	474.59	433.66	386.76	357.32	90.92	126.39
in-sample R^2	0.4271	0.0000	0.0000	0.0000	0.0000	0.0013	0.0015
in-sample Adj. R^2	0.4251	-0.0002	-0.0002	-0.0002	-0.0002	-0.0003	-0.0001
in-sample RMSE	0.8885	1.1700	1.1694	1.1659	1.1595	1.1535	1.1454
out-of-sample RMSE	**0.8363**	**1.1160**	**1.1236**	**1.1332**	**1.1439**	**1.1531**	**1.1655**
out-of-sample AR6 RMSE	**0.7076**	**0.9456**	**0.9513**	**0.9594**	**0.9687**	**0.9766**	**0.9877**
out-of-sample MA3 RMSE	**0.8295**	**0.9456**	**0.9511**	**0.9601**	**0.9710**	**0.9761**	**0.9851**

In contrast to the low number of selected predictors in anti-Asian results (state-level), anti-Hispanic prediction models (Table 5.4) include between 7 and 21 selected predictors. In previous results, lags of hate crime have been most often included in models that indicate over-fitting – for example, Table 5.1, model (1) and Table 5.2, models (1), (3), (5). However, for state-level anti-Hispanic hate crime prediction models, lags of the dependent variable are selected for all models, which may indicate the utility of past levels of hate crime in predicting anti-Hispanic hate crime on a state-level.

Table 5.4: Anti-Hispanic Results (State-level, 2004-2019)

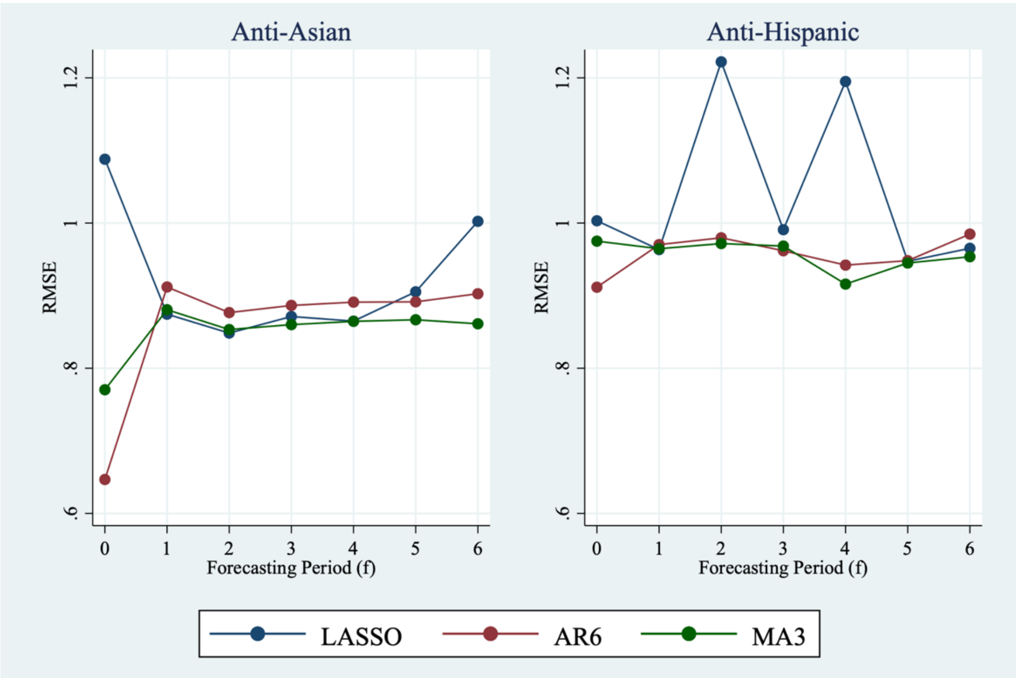
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
wetback	0.0142			-0.0149	0.0173	-0.0417	
$wetback^2$	0.0248		-0.0260	0.0135		0.0193	
wetbacks	0.0071			-0.0469	0.0059	0.0183	0.0024
$wetbacks^2$	-0.0075		-0.0005	0.0281	-0.0030		
beaner	-0.0004			-0.0152		0.0293	-0.0072
$beaner^2$				-0.0012	0.0072	-0.0309	0.0081
beaners	0.0321	-0.0035		-0.0191	0.0310	-0.0442	-0.0173
$beaners^2$				-0.0067	-0.0245	0.0456	0.0221
spic	0.0004		-0.0019			0.0314	
$spic^2$	0.0179		-0.0030		-0.0217	-0.0116	0.0254
spics	0.0147			0.0029	-0.0064		
$spics^2$			0.0185	-0.0560		0.0046	
Total Employment				0.0283			
Employment (%)				0.0099	-0.0003	0.0786	-0.0156
Total Labour Force Participation							
Labour Force Participation (%)	0.0003		0.0092			-0.0813	
Total Unemployment	-0.0327			-0.0168	-0.0045	-0.0014	-0.0005
Unemployment (%)	-0.0025		-0.0052		-0.0035	0.0594	-0.0048
Total Asian Population							
Asian Population (%)	0.0118		0.0013	0.0129	0.0013	0.0006	
Total Hispanic Population	0.0045					-0.0052	
Hispanic Population (%)							
Total Population							
Anti-Hispanic Hate Crime (Lag 1)	-0.7858	0.0287	-0.0989	0.0865	-0.0656	0.0169	-0.0739
Anti-Hispanic Hate Crime (Lag 2)	-0.5650	-0.0721	0.0134	-0.0016	-0.0271	-0.0637	0.0125
Anti-Hispanic Hate Crime (Lag 3)	-0.4906	0.0315	-0.0428	0.0067	-0.0887	0.0190	
Anti-Hispanic Hate Crime (Lag 4)	-0.2995	-0.0218	-0.0169	-0.0614			0.0104
Anti-Hispanic Hate Crime (Lag 5)	-0.2029		-0.0668	0.0066	-0.0052	0.0068	

Anti-Hispanic Hate Crime (Lag 6)	-0.0796	-0.0525		-0.0048	0.0105	-0.0155	0.0319
Constant	-0.0036	-0.0024	-0.0005	0.0031	-0.0037	0.0002	0.0010
Selected Predictors	20	7	14	20	17	21	14
Selected Optimal λ	17.6	138.9	54.35	17.98	38.53	18.64	43.76
in-sample R^2	0.4024	0.0215	0.0220	0.0189	0.0126	0.0119	0.0104
in-sample Adj. R^2	0.4000	0.0202	0.0194	0.0150	0.0091	0.0077	0.0078
in-sample RMSE	0.9085	1.1612	1.1580	1.1590	1.1654	1.1663	1.1668
out-of-sample RMSE	**1.0233**	**1.1388**	**1.1473**	**1.1649**	**1.1548**	**1.1667**	**1.1771**
out-of-sample AR6 RMSE	**0.8627**	**0.9636**	**0.9704**	**0.9837**	**0.9768**	**0.9844**	**0.9951**
out-of-sample MA3 RMSE	**0.9198**	**0.9645**	**0.9628**	**0.9851**	**0.9751**	**0.9799**	**0.9889**

5.3 Model Comparison

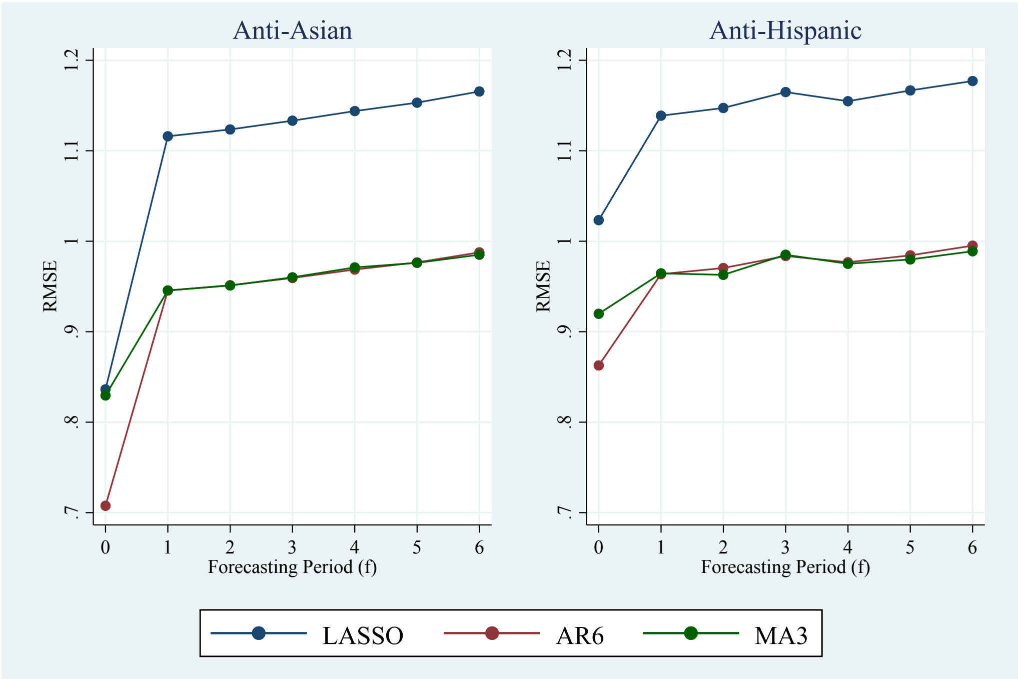
In order to evaluate the out-of-sample prediction performance the LASSO selected prediction models (LASSO) are compared to autoregressive processes of order 6 (AR6) and moving average processes of order 3 (MA3) – simpler and less computationally intensive prediction models.

Figure 5.1: Out-of-Sample Prediction Performance for LASSO vs. AR(6) vs. MA(3) (National-level, 2004-2019)



In the majority of cases, the LASSO models are outperformed by the AR(6) and MA(3) benchmark models. Figure 5.1 and 5.2 illustrate this clearly, showing the national-level and state-level out-of-sample performance, respectively, for all selected anti-Asian and anti-Hispanic prediction models in comparison to the simpler benchmark models. Only national-level anti-Asian LASSO models which predict hate crime one $(f = 1)$ and two $(f = 2)$ periods in the future, outperform both benchmarks. The difference in out-of-sample prediction error (RMSE) between LASSO prediction models and benchmark models, AR6 and MA3, is much larger on the state-level than national-level – in both anti-Asian and -Hispanic hate crime prediction models.

Figure 5.2: Out-of-Sample Prediction Performance for LASSO vs. AR(6) vs. MA(3) (State-level, 2004-2019)



6 ROBUSTNESS CHECKS

The potential bias ensuing from the applied methodology is addressed in Section 6.1, whilst Section 6.2 focuses on accounting for structural breaks to improve prediction performance.

6.1 Methodology Bias

As previously mentioned in Section 4.3, the implementation of cross-validation in a time-series setting for a non-autoregressive model can lead to a downward bias due to the violation of independence between training and test sets. To address this, Table 6.1 and 6.2 present results of purely autoregressive processes of order six (LAR6) run using cross-validation and LASSO. LAR6 results offer an unbiased benchmark with which to compare and analyse how downward biased main results fair.

The large CV-selected penalisation coefficients (λ) cause, for the majority of anti-Asian hate crime prediction models (Table 6.1), the exclusion of lags for models (2)-(7). However, due to the much lower λ value, all 6 lags of anti-Asian hate crime are included in model (1) to predict hate crime in the present ($f = 0$). In contrast, multiple lags are included in each prediction model for anti-Hispanic hate crime (Table 6.2) as the penalisation co-efficient varies between 0.54 and 7.24.

Table 6.1: Anti-Asian LASSO AR6 Results (National-level, 2004-2019)

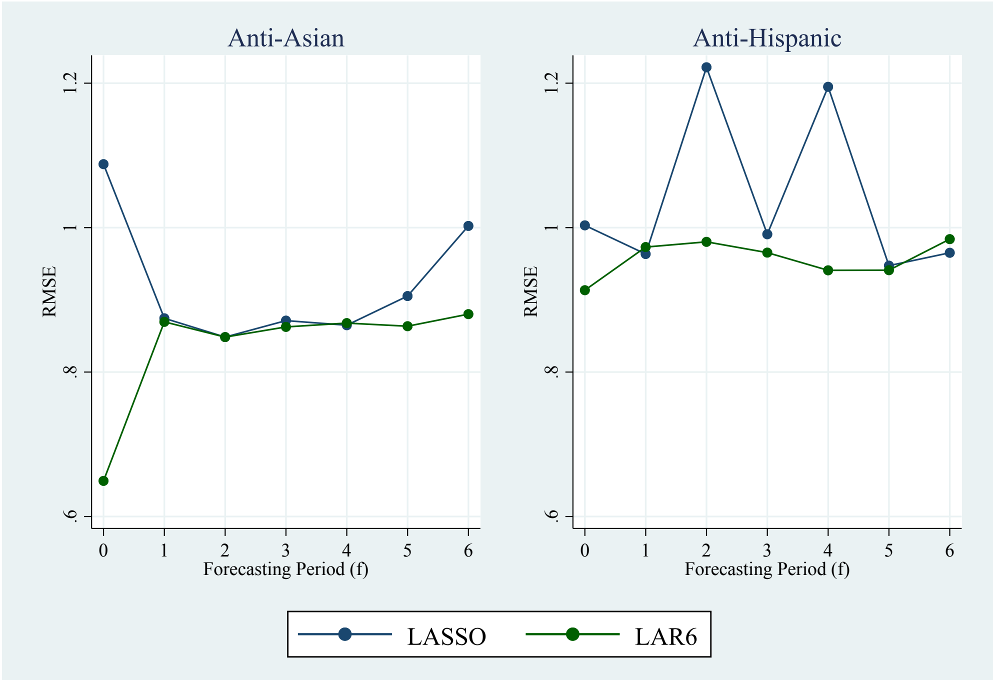
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Anti-Hispanic Hate Crime (Lag 1)	-0.7328						
Anti-Hispanic Hate Crime (Lag 2)	-0.6144						
Anti-Hispanic Hate Crime (Lag 3)	-0.4291						
Anti-Hispanic Hate Crime (Lag 4)	-0.4054						
Anti-Hispanic Hate Crime (Lag 5)	-0.3179						
Anti-Hispanic Hate Crime (Lag 6)	-0.2798						0.0495
Constant	-0.0294	-0.0094	0.0045	-0.0094	-0.0141	0.0057	-0.0095
Selected Predictors	6	0	0	0	0	0	1
Selected Optimal λ	0.87	42.69	39.33	33.17	33.59	30.04	32.98
out-of-sample LAR6 RMSE	**0.6493**	**0.8695**	**0.8485**	**0.8625**	**0.8678**	**0.8635**	**0.8802**
out-of-sample Model RMSE	**1.0879**	**0.8745**	**0.8485**	**0.8713**	**0.8649**	**0.9053**	**1.0024**

Table 6.2: Anti-Hispanic LASSO AR6 Results (National-level, 2004-2019)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Anti-Hispanic Hate Crime (Lag 1)	-0.4085	-0.0089	-0.0764	0.0158	-0.2024		-0.1288
Anti-Hispanic Hate Crime (Lag 2)	-0.1377	-0.0801		-0.2111	-0.0709	-0.1504	0.0014
Anti-Hispanic Hate Crime (Lag 3)	-0.1129		-0.2050	-0.1019	-0.1694		-0.0408
Anti-Hispanic Hate Crime (Lag 4)		-0.2018	-0.0878	-0.1988		-0.0505	
Anti-Hispanic Hate Crime (Lag 5)	-0.1455	-0.0798	-0.1713	-0.0320	-0.0419		0.0297
Anti-Hispanic Hate Crime (Lag 6)	-0.0482	-0.1670	-0.0117	-0.0514	-0.0201	-0.0023	0.1520
Constant	-0.0065	-0.0125	-0.0059	0.0045	-0.0095	-0.0031	0.0023
Selected Predictors	5	5	5	6	5	3	5
Selected Optimal λ	7.24	3.53	2.9	0.54	4.89	4.45	2.95
out-of-sample LAR6 RMSE	**0.9133**	**0.9731**	**0.9803**	**0.9654**	**0.9409**	**0.9411**	**0.9841**
out-of-sample Model RMSE	**1.0032**	**0.9635**	**1.2221**	**0.9908**	**1.1950**	**0.9474**	**0.9652**

Not surprisingly, Figure 6.1 reveals that on average optimistically biased LASSO prediction results are outperformed by an unbiased LAR6. Although this approach doesn't control for the bias itself, it does provide perspective. Namely, if main out-of-sample results accounted for the bias, they would fair even worse in comparison with LAR6 than shown in Figure 6.1 because of their already poor prediction performance. For this reason, there is little need to account for the methodology bias further.

Figure 6.1: RMSE Out-of-Sample model vs. LAR6 for Anti-Asian & -Hispanic Hate Crime (National-level, 2004-2019)



6.2 Structural Break: Donald J. Trump

A structural break is an unpredictable change over time, often over external events such as economic or political events (Antoch et al., 2018), in the parameters of a regression model, in this case the LASSO selected model, which can adversely impact the performance of a forecasting model (Hashem Pesaran et al., 2006; Maheu & Gordon, 2008). Bai and Perron (1998) developed following methods to test for and estimate multiple structural breaks, where $s = 1, \dots, s^*$:

Hypothesis 1:	H_0: no structural break	vs.	H_1: s number of breaks
Hypothesis 2:	H_0: no structural break	vs.	H_1: $1 \leq s \leq s^*$ breaks
Hypothesis 3:	H_0: s number of breaks	vs.	H_1: $s + 1$ number of breaks

Following testing for multiple breaks through Hypothesis 2, Hypothesis 3 helps to fine-tune the actual number of breaks present. Table 6.3 shows the results of testing Hypothesis 2 and 3 for structural breaks in anti-Asian and -Hispanic hate crime. Both hate crimes significantly reject Hypothesis 2's null hypothesis, indicating that the number of breaks ranges from 1 to 5. As all Hypothesis 3 tests for anti-Asian hate crimes fail to reject the null hypothesis, it can be assumed that only one break exists. For anti-Hispanic hate crime, the null hypothesis that only one structural break exists is rejected whilst all other Hypothesis 3 tests fail to reject the null hypothesis – implying two breaks exist.

Table 6.3: Results of Testing for Structural Breaks

	Hypothesis	H_0	H_1	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Anti-Asian Hate Crime	2	No breaks	$1 \leq s \leq 5$ breaks	58.37***	12.37	8.88	7.46
	3	1 break	2 breaks	6.96	13.89	10.13	8.51
		2 break	3 breaks	2.91	14.80	11.14	9.41
		3 break	4 breaks	2.26	15.28	11.83	10.04
		4 break	5 breaks	1.81	15.76	12.25	10.58
Anti-Hispanic Hate Crime	2	No breaks	$1 \leq s \leq 5$ breaks	56.97***	12.37	8.88	7.46
	3	1 breaks	2 breaks	19.49***	13.89	10.13	8.51
		2 breaks	3 breaks	3.59	14.80	11.14	9.41
		3 breaks	4 breaks	2.71	15.28	11.83	10.04
		4 breaks	5 breaks	2.19	15.76	12.25	10.58

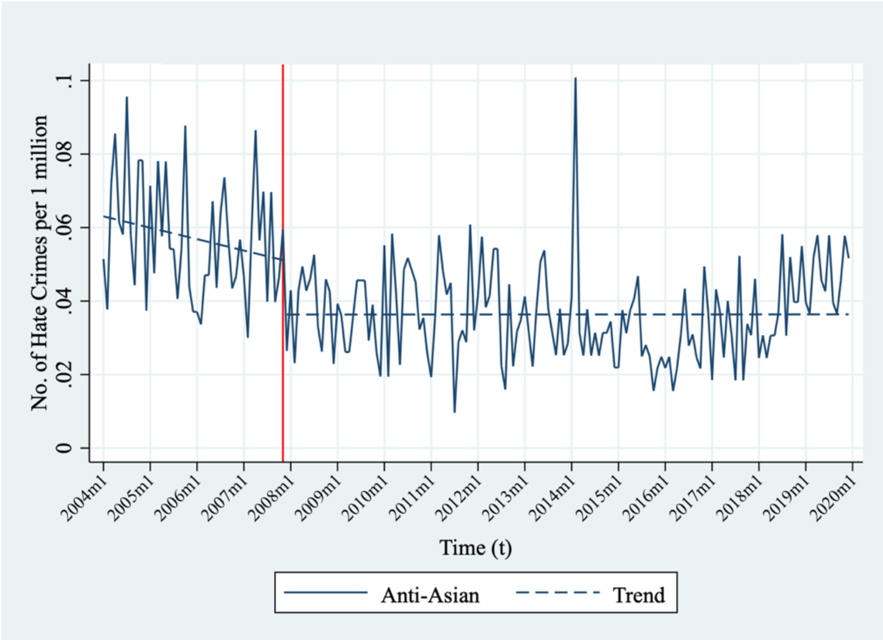
*** - the test statistic is significant at the 1% critical value.

In addition to testing for structural breaks, Bai and Pierron (1998) method's allow for estimating the time and confidence intervals of structural intervals – results shown in Table 6.4. The anti-Asian (AS) breakpoint AS.1 is estimated to be at November 2007, anti-Hispanic (HI) breakpoints HI.1 at October 2010 and HI.2 at October 2016. Figure 6.2 and 6.3 help to visualise the estimated breakpoints for anti-Asian and -Hispanic hate crimes, respectively.
Theoretically, accounting for a structural break ought to improve prediction accuracy. To be able to account for one of the three discovered breaks, it is necessary to identify and then account for an external event which impacts the corresponding hate crimes. In and around November 2007 there are no significant political or economic events that would suggest a structural change in anti-Asian hate crime – especially an event which stabilises the trend in hate crimes as seen in Figure 6.2 between November 2007 and December 2019.

Table 6.4: Structural Breakpoint Estimation and Confidence Intervals

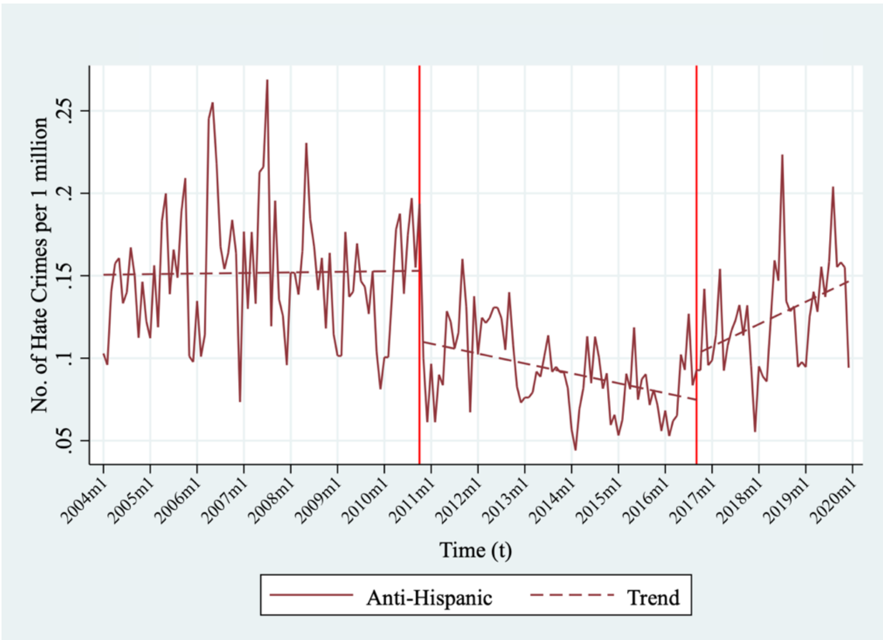
	#	Estimated Breakpoints	95% Confidence Interval	
Anti-Asian Hate Crime	(AS.1)	November 2007	October 2007	November 2007
Anti-Hispanic Hate Crime	(HI.1)	October 2010	September 2010	October 2010
	(HI.2)	October 2016	September 2016	October 2016

Figure 6.2: Anti-Asian Hate Crime with Structural Breaks (National-level, 2004-2019)



Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

Figure 6.3: Anti-Hispanic Hate Crime with Structural Breaks (National-level, 2004-2019)



Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

Similarly, there is no event in and around October 2010 (Figure 6.3) which suggests a declining trend in anti-Hispanic crimes for the following 6 years. However, looking at October 2016 (Figure 6.3), the running for and election of Donald J. Trump as U.S. president is the most likely major U.S. political event to influence the trend of anti-Hispanic hate crimes. It is hypothesised that accounting for Trump's presidential campaign and eventual election (structural break) in model selection and training would improve prediction accuracy in comparison to benchmarks. Figure 6.4 puts the October 2016 breakpoint into perspective of both anti-Asian and -Hispanic hate crime. Anti-Asian hate crime is graphically included merely to also highlight a trend change in anti-Asian hate crime that overlaps with that of anti-Hispanic hate crime.

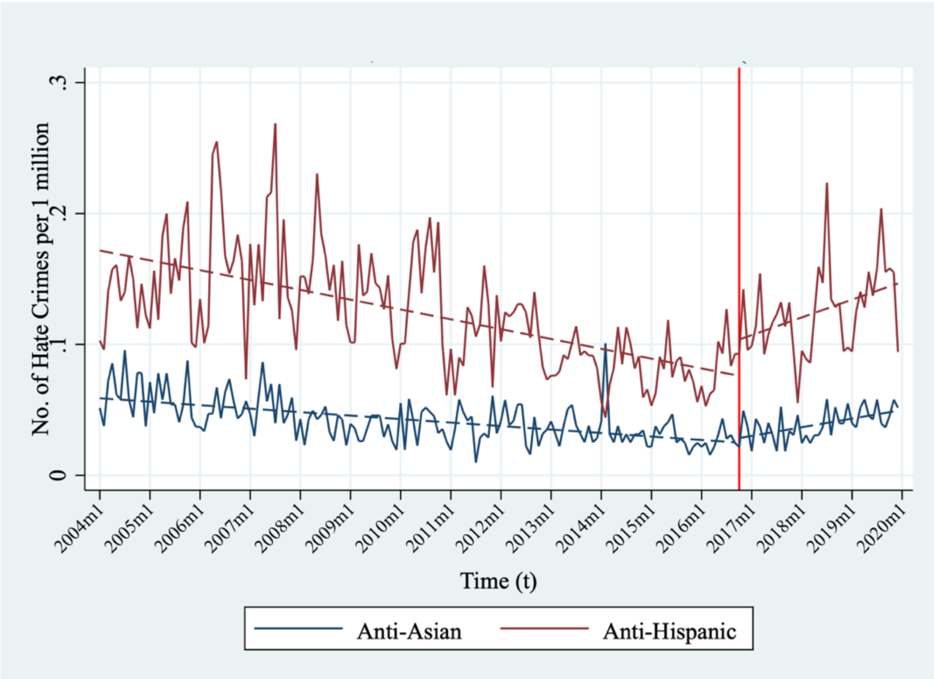
Research shows increased levels of discrimination Hispanics endured during Trump's presidential campaign, election and tenure as U.S. president (Lopez et al., 2018, pp. 22-24; Canizales & Vallejo, 2021; Rushin & Edwards, 2018). To verify Trump's presidential campaign and election as U.S. president as a point of structural the out-of-sample RMSE of the models trained on data before Trump's presidential campaign (hereafter referred as Pre-Trump), 2004 to 2015, and during and after his campaign (hereafter referred as Post-Trump), 2006 to 2017, are compared with their corresponding benchmarks. The same methodology as implemented for the national-level results is used with additional Google Trends data on Trump-related terms, shown in Table 6.5 - where the included Trump-related terms, based upon racial rhetoric used by Trump, are sourced from news articles from Time (Reilly, 2016) and Vox (Lopez, 2020).

Table 6.5: Trump-related Anti-Hispanic Rhetoric & Selected Google Trends Search Terms

Trump-related (Reilly, 2016; Lopez, 2020)	**Selected Search Terms**
Mexican rapists	rapist, rapists
Mexican immigrants	immigrant, immigrants
Mexican wall	mexican wall
Mexican(s)	mexican, mexicans

For readability, Tables 6.6 and 6.7 exclude unselected variables, which in both cases are all demographic variables, additional Google Trends terms (listed in Table 6.5) and lags of anti-Hispanic hate crime[12]. The non-selection of these variables can be caused due to either the high correlation with other selected variables or their poor ability to predict the dependent variable. The latter is the more likely reason for exclusion as Appendix: Table A.4 shows the lack of high correlation between the included and excluded variables.

Figure 6.4: Hispanic Hate Crime Structural Break of October 2016 (National-level, 2004-2019)



Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

No pre-Trump prediction model (Table 6.6) is able to explain more than 10.17% of in-sample variation of future anti-Hispanic hate crime. Although model (4) has the most included racial slurs (13), it also has the highest out-of-sample RMSE of 1.0389 indicating overfitting on the training sample. The extremely high levels of λ selected by CV explains the lack of predictors in models (2), (3), (5), (6) and (7). Table 6.7 shows the number of selected variables and the in-sample adjusted R^2 for post-Trump models (3)-(7) ranging between 1 to 19 and 1.76% to 8.87%, respectively. In contrast, Model (2) which has no predictors and has an out-of-sample prediction error of 1.0236, explains 0.00% in-sample variation of anti-Hispanic hate crime. High levels of λ are also prevalent in the Post-Trump models (2), (5) and (7). Models

[12] The mentioned demographic variables are those also included in the running of LASSO on the national-level, seen in Table 5.1.

(1) and (3) appear to suffer from overfitting of in-sample data because of the numerous included predictors and much lower in-sample prediction error in relation to the out-of-sample error.

Table 6.6: Anti-Hispanic Results Pre-Trump Presidential Campaign (National-level, 2004-2019)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
wetback	-0.0008			0.0023			
wetback²	0.0002			-0.0002			
wetbacks				0.0050			
wetbacks²	0.0001				-0.0001		
beaner							
beaner²	0.0001				0.0001		
beaners				-0.0161			
beaners²	0.0001			0.0001			
chopa				0.0031			
chopa²				0.0001			
chopas				-0.0006			
chopas²	0.0001						
chopita							
chopita²	-0.0001						
pepper belly							
pepper belly²	0.0002			-0.0001			
pepper bellys				0.0029			
pepper bellys²	-0.0001	0.0001		-0.0002			
taco bender							
taco bender²							-0.0001
spic				-0.0018			
spic²							0.0001
spics							
spics²				-0.0001			
Constant	-0.1433	-0.0557	-0.0096	0.0879	-0.0602	0.0051	0.0198
Selected Predictors	9	1	0	13	2	0	2
Selected Optimal λ	580.01	36,686.78	37,391.17	95.33	17,880.94	46,398.37	14,345.04
in-sample R^2	0.2190	0.0999	0.0280	0.2338	0.0705	0.0409	0.1447
in-sample Adj. R^2	0.1294	0.0510	0.0061	0.1017	0.0200	0.0192	0.0982
in-sample RMSE	0.8935	0.9580	0.9955	0.8814	0.9682	0.9860	0.9312
out-of-sample RMSE	**0.9252**	**0.9319**	**0.9274**	**1.0389**	**0.9710**	**0.9474**	**0.9868**
out-of-sample AR6 RMSE	**0.8417**	**0.8969**	**0.9045**	**0.9013**	**0.9183**	**0.9275**	**0.9584**
out-of-sample MA3 RMSE	**0.8933**	**0.9082**	**0.9106**	**0.9215**	**0.8995**	**0.9336**	**0.9355**

Out of a potential 24 anti-Hispanic racial slur variables, 19 are selected for at least one pre-Trump prediction model (Table 6.6), whereas 20 variables are included in a minimum of one post-Trump prediction model (Table 6.7). The selection frequency of anti-Hispanic racial slurs relative to Trump-related terms and demographic variables would suggest their importance in predicting anti-Hispanic hate crime. However, although many racial slurs are selected, comparing the prediction models out-of-sample performance with that of the AR6 and MA3 benchmarks shows that these racial slurs belong to prediction models are unable to outperform their simpler benchmark models[13]. Figure 6.4 shows both models failing to outperform their benchmarks with only the post-Trump model (1) outperforming both AR6 and MA3 benchmarks.

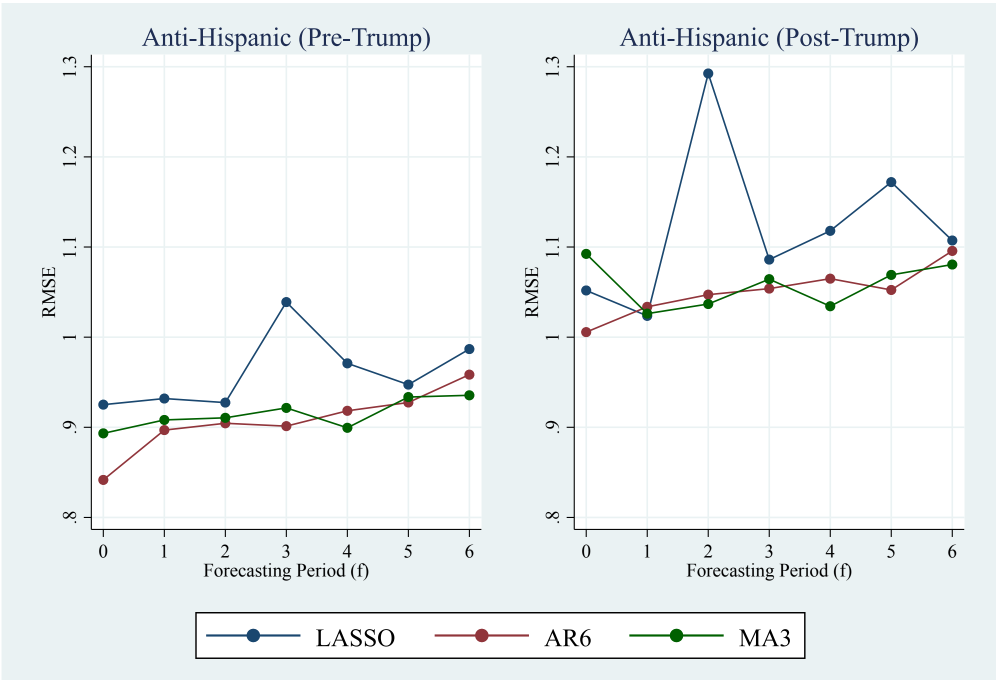
Table 6.7: Anti-Hispanic Results Post-Trump Presidential Campaign (National-level, 2004-2019)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
wetback	0.2822		-0.2900	-0.1147		0.9700	
wetback²	-0.1334		0.1273				
wetbacks	0.2353		-0.5685				
wetbacks²	0.9250		-0.2300	0.8190	-0.5460	0.3440	
beaner							
beaner²	0.6210		-0.1390	-0.3770	0.8450	0.1530	
beaners	0.6920		0.9881				
beaners²	0.8900		-0.4560	-0.5880	0.3770	-0.1930	
chopa	-0.3480		-0.5436				
chopa²	-0.3940		0.2800	0.1436		-0.2700	
chopas	0.1433		-0.1920	-0.2120	-0.2230	0.9640	
chopas²	-0.3751		-0.8667				
chopita	-0.5675						
chopita²	-0.4300		-0.6960	-0.2790		-0.9990	
pepper belly							
pepper belly²	0.8000		0.3700	-0.8620		-0.3181	
pepper bellys			-0.7895				
pepper bellys²	-0.9200		0.8120	-0.2150		0.3690	
taco bender							
taco bender²	0.5200		-0.6120	-0.4740		0.6970	
spic	0.3730		0.3870	-0.1780	0.1770	-0.5760	0.6570
spic²							
spics	0.3532		0.7912				
spics²	-0.1790		-0.8100	-0.5370	0.2990	0.2400	
Constant	-0.0940	-0.0035	0.0926	-0.0229	-0.0427	-0.0743	0.0141
Selected Predictors	14	0	13	7	2	7	1
Selected Optimal λ	164.98	138,013.60	110.34	912.88	22,033.31	785.34	115,424.80
in-sample R^2	0.2548	0.0000	0.1448	0.1404	0.0605	0.1237	0.1001
in-sample Adj. R^2	0.1483	-0.0063	0.0226	0.0644	0.0176	0.0462	0.0887
in-sample RMSE	0.8711	1.0116	0.9350	0.9348	0.9783	0.9469	0.9598
out-of-sample RMSE	**1.0518**	**1.0236**	**1.2926**	**1.0860**	**1.1180**	**1.1720**	**1.1073**
out-of-sample AR6 RMSE	**1.0056**	**1.0337**	**1.0472**	**1.0539**	**1.0649**	**1.0524**	**1.0957**
out-of-sample MA3 RMSE	**1.0924**	**1.0261**	**1.0368**	**1.0644**	**1.0343**	**1.0691**	**1.0806**

[13] As each model draws upon different time periods of the dataset, it is important to measure out-of-sample performance by comparing each model against its relevant AR6 and MA3 benchmarks – that have been measured over the same sample – rather than the out-of-sample performance with each other.

The results shown in Table 6.6 and 6.7 imply that even the accounting for the HS.2 breakpoint in the model, through additional data and training of models on data before and after the structural break, fails to improve the prediction accuracy above its benchmarks. This failure, in combination with the poor prediction performance of main results, is most likely either due to the use of cross-validation and LASSO to create prediction models or search popularity of racial slurs' inadequate power to accurately predict.

Figure 6.4: Pre-& Post-Trump Out-of-Sample Performance Model vs. AR6 vs. MA3 (National-level)



7 DISCUSSION

Section 7.1 will address the failure of LASSO prediction models to outperform their benchmarks on the national- and state-level. Thereafter, the discussion of this paper's findings will be reviewed in the scope of relevant literature and the broader field of crime prediction within Section 7.2.

7.1 Discussion of Results

As discussed in Section 3.4.2, stationary data provides the foundation for improved prediction performance of forecasting models. Therefore, the partial stationarity of the national-level (64% of the dataset) may in fact be limiting the predictive accuracy of the LASSO models.

However, this could be disputed based upon the state-level results. Despite a completely stationary set of state-level variables, state-level prediction models perform far worse against their benchmarks (AR6 and MA3 models) than national-level prediction models where 36% of data is non-stationary. This either indicates that non-stationarity may be less a limitation than expected or that state-level LASSO hate crime prediction models systematically struggle predict accurately more than those on the national-level. As the former, calculating the limitation of not using a completely stationary time-series, is outside the objective of this research, reasons for the latter are discussed.

6.28% and 3.66% of national-level observation values for anti-Asian and anti-Hispanic hate crime, respectively, are 0. Whereas, 69.93% of anti-Asian and 49.15% anti-Hispanic hate crime observation values are 0 on the state-level. This may partly explain why the simpler AR6 and MA3 models significantly outperform the complex LASSO prediction models more on the state-level than national-level, as there is little to no change in hate crime against Asians and Hispanics from period to period on the state-level. For example, a simple model which predicted no change in a state's hate crime against Asians and Hispanics would be correct in approximately 70% and 49% of the cases, respectively. Therefore, state-level results may be less reliable and must be interpreted with caution.

The fact that many of the LASSO models are overfitted or do not include racial slur search popularity variables, especially those with the best out-of-sample prediction performance, suggests the limitation of using racial slur Google search popularity to predict hate crime well. This is reinforced by the underperformance of LASSO prediction models, relative to benchmarks, in the anti-Hispanic state-level results (Table 5.4) where on average 7 racial slur search popularity predictors are included in a prediction model. Furthermore, out of the only two LASSO prediction models (Model (2), Table 5.1) to outperform their benchmarks, one includes one racial slur search popularity predictor, whilst the other (Model (3), Table 5.1) doesn't select a single racial slur search popularity variable.

In summary, it can be strongly argued that Google Trends data, based on the search popularity of racial slurs used in this paper, poorly predicts anti-Asian and -Hispanic hate crime. However, it cannot be completely ruled out that these discouraging results may in part be caused due to the methodological bias, the implementation of CV and LASSO, or a lack of sufficient breadth and depth in racial slur search popularity data. With an eye towards future research in the field of hate crime prediction using Google Trends data, employing more common methods of crime prediction, such as spatial analysis (Caplan et al., 2011; Mohler, 2014; Chainey et al., 2018; Liang et al., 2022; Jendryke et al., 2021) or machine learning methods (Reier Forradellas et al., 2020; Wang, 2021) and expanding the collection of search popularity data across a plethora of racial slurs, or racial antagonistic language, may assist in building upon the results presented in this paper.

7.2 Findings and Relevant Literature

Google Trends data has been used across multiple fields of research for the purpose of prediction. In the field of health, Morsy et al. (2018), Lu et al. (2019) and Venkatesh and Gandhi (2020) use Google Trends data to predict the level of zika virus, avian influenza and COVID-19 cases, respectively. In the political realm, Prado-Román et al. (2020) successfully predicted presidential election winners between 2004 and 2020 using search popularity data from Google. Hamid and Heiden (2015), Ahmed et al. (2017) and Petropoulos et al. (2021) forecast volatilities in financial markets using Google Trends. Although Google Trends data has been used to predict "meth"-related crimes (Gamma et al., 2016) and a large variety of crime in Mexico City (Piña-García and Ramírez-Ramírez, 2019)[14], it is believed not yet to have been used to predict anti-Asian hate crime. Therefore, this paper, withstanding its limited results, is the first to extend the above stated literature by exploring the use of Google Trends data as a predictor for hate crime – specifically anti-Asian and -Hispanic hate crime.

Much research has endeavoured to predict (hate) crime. Williams et al. (2019) predict hate crimes through the use of online anti-Black and anti-Muslim hate speech data collected from Twitter. Caplan et al. (2011), Mohler (2014), Chainey et al. (2018), Jendryke et al. (2021) and Liang et al. (2022) spatially forecast (hate) crime ranging from gun-related crimes to burglaries. However, to-date there is no research available on prediction models which aim to predict true levels of hate crime (the sum of un- and reported hate crime incidents). Beyond its main findings, this paper observes trend disparity in willingness to report experienced hate crime across Asian and Hispanic groups. If this observation also holds for other types of crime, the possibility exists for others to go beyond existing (hate) crime prediction research by exploring forecasting models which forecast true levels of (hate) crimes by accounting for ethnic groups' differing trends in willingness to report their experienced (hate) crime. Approaches may include the use of survey data, such as the annual nationally representative National Crime Victimisation Survey (NCVS) conducted by the Bureau of Justice Statistics

[14] Work from Piña-García and Ramírez-Ramírez (2019) covers robbery of passer-by's, of business property, on public transportation (incl. taxi), to carrier and deliver person, theft of motor vehicles, card fraud, homicide, rape, burglary and firearm injuries.

(BJS, n.d.), which gives an indication of the level of unreported hate crime in relation to those reported.

An additional finding of this research are structural breaks which help identify systematic changes in anti-Asian and -Hispanic hate crime trends. Although accounting for the structural break, Trump's presidential campaign and election as U.S. president, ultimately didn't improve hate crime prediction power, it highlights a shift in hate crime towards Hispanics. Such a shift suggests the potential for national events or public figures in a position of authority to impact the degree of hate crime towards specific ethnic groups. In the specific case of Trump's presidential campaign and election, research validates the relationship between Trump and rising anti-Hispanic hate crime (Lopez et al., 2018, pp. 22-24; Canizales & Vallejo, 2021; Rushin & Edwards, 2018). It is suggested that further research examines the impact of structural breaks resulting from national events or public figures on hate crime, in other settings. This may include evaluating the possible correlation between anti-Asian hate crime and Trump during COVID-19 (2020-2022) once data is available for the former, or the number of hate crimes committed against people of Middle-Eastern origin resulting from 9/11.

8 CONCLUSION

This paper attempted to explore Google Trends' power to predict anti-Asian hate crime through prediction models constructed by LASSO. Based on the analysis of findings presented, there is no evidence that publicly available Google Trends data can predict either anti-Asian or -Hispanic hate crime without large prediction error. Consequently, it can be proposed that Google Trends data poorly predicts hate crime because it doesn't proxy well for racial bias. Future research aimed at identifying and predicting rises in hate crime may be better suited using more common prediction techniques such as spatial analysis (Caplan et al., 2011; Mohler, 2014; Chainey et al., 2018; Jendryke et al., 2021; Liang et al., 2022) or machine learning (Reier Forradellas et al., 2020; Wang, 2021) as LASSO was found to be a less suited method.

REFERENCES

Ahmed, F., Raheela, D., Saman, D., & Muzammil, M. (2017). Financial Market Prediction using Google Trends. *International Journal of Advanced Computer Science and Applications*, *8*(7). https://doi.org/10.14569/ijacsa.2017.080752

Ahmed, S., Hsueh-Hua, V. C., & Chib, A. I. (2021). Xenophobia in the Time of a Pandemic: Social Media Use, Stereotypes, and Prejudice against Immigrants during the COVID-19 Crisis. *International Journal of Public Opinion Research*. Published. https://doi.org/10.1093/ijpor/edab014

Ahrens, A., Hansen, C. B., & Schaffer, M. E. (2020). lassopack: Model selection and prediction with regularized regression in Stata. *The Stata Journal: Promoting Communications on Statistics and Stata*, *20*(1), 176–235. https://doi.org/10.1177/1536867x20909697

Antoch, J., Hanousek, J., Horváth, L., Hušková, M., & Wang, S. (2018). Structural breaks in panel data: Large number of panels and short length time series. *Econometric Reviews*, *38*(7), 828–855. https://doi.org/10.1080/07474938.2018.1454378

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*. https://doi.org/10.1214/09-ss054

Awan, I., & Zempi, I. (2015). 'I will Blow your face off'—Virtual and Physical World Anti-Muslim Hate Crime. *British Journal of Criminology*, azv122. https://doi.org/10.1093/bjc/azv122

Awan, I., & Zempi, I. (2016). The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts. *Aggression and Violent Behavior*, *27*, 1–8. https://doi.org/10.1016/j.avb.2016.02.001

Bai, J., & Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, *66*(1), 47. https://doi.org/10.2307/2998540

Barnes, A., & Ephross, P. H. (1994). The Impact of Hate Violence on Victims: Emotional and Behavioral Responses to Attacks. *Social Work*, *39*(3), 247–251. https://doi.org/10.1093/sw/39.3.247

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70–83. https://doi.org/10.1016/j.csda.2017.11.003

Benier, K. (2017). The harms of hate: Comparing the neighbouring practices and interactions of hate crime victims, non-hate crime victims and non-victims. *International Review of Victimology*, *23*(2), 179–201. https://doi.org/10.1177/0269758017693087

Benjamin, E. (2020). Trump, the Coronavirus Pandemic, Asian American Xenophobia, and Humanistic Psychology. *Journal of Humanistic Psychology*, *61*(2), 244–259. https://doi.org/10.1177/0022167820979650

Bennett Moses, L., & Chan, J. (2016). Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society*, *28*(7), 806–822. https://doi.org/10.1080/10439463.2016.1253695

Berinsky, A. J. (1999). The Two Faces of Public Opinion. *American Journal of Political Science*, *43*(4), 1209. https://doi.org/10.2307/2991824

Borell, K. (2015). When Is the Time to Hate? A Research Review on the Impact of Dramatic Events on Islamophobia and Islamophobic Hate Crimes in Europe. *Islam and Christian–Muslim Relations*, *26*(4), 409–421. https://doi.org/10.1080/09596410.2015.1067063

Bureau of Justice Statistics (BJS). (n.d.). *Bureau of Justice Statistics (BJS) - National Crime Victimization Survey (NCVS)*. Data Collection: National Crime Victimization Survey (NCVS). Retrieved May 13, 2021, from https://www.bjs.gov/index.cfm?ty=dcdetail&iid=245

Burman, P., Chow, E., & Nolan, D. (1994). A cross-validatory method for dependent data. *Biometrika*, *81*(2), 351–358. https://doi.org/10.1093/biomet/81.2.351

Canizales, S. L., & Vallejo, J. A. (2021). Latinos & Racism in the Trump Era. *Daedalus*, *150*(2), 150–164. https://doi.org/10.1162/daed_a_01852

Chang, K. (2009). Circulating Race and Empire: Transnational Labor Activism and the Politics of Anti-Asian Agitation in the Anglo-American Pacific World, 1880–1910. *Journal of American History*, *96*(3), 678–701. https://doi.org/10.1093/jahist/96.3.678

Chen, J. A., Zhang, E., & Liu, C. H. (2020). Potential Impact of COVID-19–Related Racial Discrimination on the Health of Asian Americans. *American Journal of Public Health*, *110*(11), 1624–1627. https://doi.org/10.2105/ajph.2020.305858

Chiu, A. (2020, March 20). *Trump has no qualms about calling coronavirus the 'Chinese Virus.' That's a dangerous attitude, experts say.* Washington Post. https://www.washingtonpost.com/nation/2020/03/20/coronavirus-trump-chinese-virus/

Cohen, M. A., & Miller, T. R. (1998). The Cost of Mental Health Care for Victims of Crime. *Journal of Interpersonal Violence*, *13*(1), 93–110. https://doi.org/10.1177/088626098013001006

Cornaglia, F., Feldman, N. E., & Leigh, A. (2014). Crime and Mental Well-Being. *Journal of Human Resources*, *49*(1), 110–140. https://doi.org/10.3368/jhr.49.1.110

Costello, M. B. (2016, April 13). *The Trump Effect: The Impact of the Presidential Campaign on Our*. Southern Poverty Law Center. Retrieved April 29, 2022, from https://www.splcenter.org/20160413/trump-effect-impact-presidential-campaign-our-nations-schools

Croom, A. M. (2014). Spanish slurs and stereotypes for Mexican-Americans in the USA: A context-sensitive account of derogation and appropriation. *Pragmática Sociocultural / Sociocultural Pragmatics*, *2*(2), 145–179. https://doi.org/10.1515/soprag-2014-0007

Croom, A. M. (2018). Asian slurs and stereotypes in the USA. *Pragmatics and Society*, *9*(4), 495–517. https://doi.org/10.1075/ps.14027.cro

Croucher, S. M., Nguyen, T., & Rahmani, D. (2020). Prejudice Toward Asian Americans in the Covid-19 Pandemic: The Effects of Social Media Use in the United States. *Frontiers in Communication*, *5*. https://doi.org/10.3389/fcomm.2020.00039

Cummings, S., & Lambert, T. (1997). Anti-Hispanic and Anti-Asian Sentiments among African Americans. *Social Science Quarterly, 78*(2), 338-353. Retrieved May 10, 2021, from http://www.jstor.org/stable/42864341

Darling-Hammond, S., Michaels, E. K., Allen, A. M., Chae, D. H., Thomas, M. D., Nguyen, T. T., Mujahid, M. M., & Johnson, R. C. (2020). After "The China Virus" Went Viral: Racially Charged Coronavirus Coverage and Trends in Bias Against Asian Americans. *Health Education & Behavior*, *47*(6), 870–879. https://doi.org/10.1177/1090198120957949

Detotto, C., & Otranto, E. (2010). Does Crime Affect Economic Growth? *Kyklos*, *63*(3), 330–345. https://doi.org/10.1111/j.1467-6435.2010.00477.x

Disha, I., Cavendish, J. C., & King, R. D. (2011). Historical Events and Spaces of Hate: Hate Crimes against Arabs and Muslims in Post-9/11 America. *Social Problems*, *58*(1), 21–46. https://doi.org/10.1525/sp.2011.58.1.21

Duncan, D. T., Hatzenbuehler, M. L., & Johnson, R. M. (2014). Neighborhood-level LGBT hate crimes and current illicit drug use among sexual minority youth. *Drug and Alcohol Dependence*, *135*, 65–70. https://doi.org/10.1016/j.drugalcdep.2013.11.001

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, *64*(4), 813. https://doi.org/10.2307/2171846

Espiritu, A. (2004). Racial diversity and hate crime incidents. *The Social Science Journal*, *41*(2), 197–208. https://doi.org/10.1016/j.soscij.2004.01.006

Executive Order 9066 (1942, February 19). *General Records of the Unites States Government*. Record Group 11. National Archives.

Fallows, J. (2020, March 19). *2020 Time Capsule #5: The 'Chinese Virus.'* The Atlantic. https://www.theatlantic.com/notes/2020/03/2020-time-capsule-5-the-chinese-virus/608260/

Fatehkia, M., O'Brien, D., & Weber, I. (2019). Correlated impulses: Using Facebook interests to improve predictions of crime rates in urban areas. *PLOS ONE*, *14*(2). https://doi.org/10.1371/journal.pone.0211350

Federal Bureau of Investigation (FBI). (n.d.). *Hate Crime*. FBI. Retrieved May 13, 2021, from https://ucr.fbi.gov/hate-crime

Federal Bureau of Investigation (FBI). (2020, November 16). *FBI Releases 2019 Hate Crime Statistics*. https://www.fbi.gov/news/pressrel/press-releases/fbi-releases-2019-hate-crime-statistics

Federal Bureau of Investigation (FBI). (2021, April 13). *UCR Publications*. Federal Bureau of Investigation. https://www.fbi.gov/services/cjis/ucr/publications

Federal Bureau of Investigation (FBI). (2022, January 3). *Hate Crimes*. Federal Bureau of Investigation. Retrieved March 5, 2022, from https://www.fbi.gov/investigate/civil-rights/hate-crimes

Gamma, A., Schleifer, R., Weinmann, W., Buadze, A., & Liebrenz, M. (2016). Could Google Trends Be Used to Predict Methamphetamine-Related Crime? An Analysis of Search Volume Data in Switzerland, Germany, and Austria. *PLoS ONE*, *11(11): e0166566*. https://doi.org/10.1371/journal.pone.0166566

Gardell, M. (2015). What's Love Got to Do with It? Ultranationalism, Islamophobia, And Hate Crime In Sweden. *Journal of Religion and Violence*, *3*(1), 91–115. https://doi.org/10.5840/jrv20155196

Gover, A. R., Harper, S. B., & Langton, L. (2020). Anti-Asian Hate Crime During the COVID-19 Pandemic: Exploring the Reproduction of Inequality. *American Journal of Criminal Justice*, *45*(4), 647–667. https://doi.org/10.1007/s12103-020-09545-1

Gratton, B., & Merchant, E. (2013). Immigration, Repatriation, and Deportation: The Mexican-Origin Population in the United States, 1920–1950. *International Migration Review*, *47*(4), 944–975. https://doi.org/10.1111/imre.12054

Green, D. P., Strolovitch, D. Z., & Wong, J. S. (1998). Defended Neighborhoods, Integration, and Racially Motivated Crime. *American Journal of Sociology*, *104*(2), 372–403. https://doi.org/10.1086/210042

Gujarati, P. D. N. (2009). *Essentials of Econometrics* (4th ed.). McGraw-Hill.

Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*, *117*, 62–81. https://doi.org/10.1016/j.jebo.2015.06.005

Hashem Pesaran, M., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting Time Series Subject to Multiple Structural Breaks. *Review of Economic Studies*, *73*(4), 1057–1084. https://doi.org/10.1111/j.1467-937x.2006.00408.x

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)* (2nd ed.). Springer.

Hate Crime Statistics Act, 28 U.S.C. § 534 Stat. (1990). https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title34-section41305&num=0&edition=prelim#referenceintext-note

Hendry, D. F. & Pretis, F. (2016). All Change! The Implications of Non-Stationarity for Empirical Modelling, Forecasting and Policy. *Oxford Martin School Policy Paper Series, Forthcoming*, Available at SSRN: https://ssrn.com/abstract=2898761

Hughes, G. (2006). *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-speaking World*. Routledge. https://doi.org/10.4324/9781315706412

Iganski, P. (2007). Too Few Jews to Count? Police Monitoring of Hate Crime Against Jews in the United Kingdom. *American Behavioral Scientist*, *51*(2), 232–245. https://doi.org/10.1177/0002764207306055

International Monetary Fund (IMF). (2021, April). Fiscal Monitor: A Fair Shot. Washington, April. https://www.imf.org/en/Publications/FM/Issues/2021/03/29/fiscal-monitor-april-2021#Full%20Report

Ismael, T., & Measor, J. (2003). Racism and the North American Media Following 11 September: The Canadian Setting. *Arab Studies Quarterly, 25*, 101-136.

Ivandic, R., Kirchmaier, T., & Machin, S. J. (2019). Jihadi Attacks, Media and Local Hate Crime. *CEPR Discussion Paper No. DP13743*. Published. https://ssrn.com/abstract=3395180

Jendryke, M., & McClure, S. C. (2021). Spatial prediction of sparse events using a discrete global grid system; a case study of hate crimes in the USA. *International Journal of Digital Earth*, *14*(6), 789–805. https://doi.org/10.1080/17538947.2021.1886356

Jung, M. H. (2005). Outlawing "Coolies": Race, Nation, and Empire in the Age of Emancipation. *American Quarterly*, *57*(3), 677–701. https://doi.org/10.1353/aq.2005.0047

Kaplan, J. (2006). Islamophobia in America?: September 11 and Islamophobic Hate Crime1. *Terrorism and Political Violence*, *18*(1), 1–33. https://doi.org/10.1080/09546550500383209

Kehoe, J. (2020). Anti-LGBTQ Hate: An Analysis of Situational Variables. *Journal of Hate Studies*, *16*(1), 21. https://doi.org/10.33972/jhs.154

Kielinger, V., & Paterson, S. (2007). Policing Hate Crime in London. *American Behavioral Scientist*, *51*(2), 196–204. https://doi.org/10.1177/0002764207306051

King, R. D. (2007). The Context of Minority Group Threat: Race, Institutions, and Complying with Hate Crime Law. *Law & Society Review*, *41*(1), 189–224. https://doi.org/10.1111/j.1540-5893.2007.00295.x

Klikauer, T. (2018). German government investigates anti-Semitism. *Ethnic and Racial Studies*, *41*(13), 2337–2343. https://doi.org/10.1080/01419870.2018.1487568

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial Attitudes and the "New South." *The Journal of Politics*, *59*(2), 323–349. https://doi.org/10.1017/s0022381600053470

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly*, *72*(5), 847–865. https://doi.org/10.1093/poq/nfn063

Lee, E. (2007). The "Yellow Peril" and Asian Exclusion in the Americas. *Pacific Historical Review*, *76*(4), 537–562. https://doi.org/10.1525/phr.2007.76.4.537

Lee, E. (2013). The Chinatown War: Chinese Los Angeles and the Massacre of 1871. *Journal of American History*, *100*(1), 217–218. https://doi.org/10.1093/jahist/jat049

Lee, D. K. (2020). Data transformation: a focus on the interpretation. *Korean Journal of Anesthesiology*, *73*(6), 503–508. https://doi.org/10.4097/kja.20137

Levin, A., Lin, C. F., & Chu, C. S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, *108*(1), 1–24. https://doi.org/10.1016/s0304-4076(01)00098-7

Light, M., & Iceland, J. (2016). The Social Context of Racial Boundary Negotiations: Segregation, Hate Crime, and Hispanic Racial Identification in Metropolitan America. *Sociological Science*, *3*, 61–84. https://doi.org/10.15195/v3.a4

Lopez, M. H., Gonzalez-Barrera, A., & Krogstad, J. M. (2018). More Latinos have serious concerns about their place in America under Trump. *Pew Research Center*, *55*. *https://www.pewresearch.org/hispanic/wp-content/uploads/sites/5/2018/10/Pew-Research-Center_Latinos-have-Serious-Concerns-About-Their-Place-in-America_2018-10-25.pdf*

Lopez, G. (2020, August 14). *Trump's long history of racism, from the 1970s to 2020*. Vox. https://www.vox.com/2016/7/25/12270880/donald-trump-racist-racism-history

Lu, Y., Wang, S., Wang, J., Zhou, G., Zhang, Q., Zhou, X., Niu, B., Chen, Q., & Chou, K. C. (2019). An Epidemic Avian Influenza Prediction Model Based on Google Trends. *Letters in Organic Chemistry*, *16*(4), 303–310. https://doi.org/10.2174/1570178615666180724103325

Maheu, J. M., & Gordon, S. (2008). Learning, forecasting and structural breaks. *Journal of Applied Econometrics*, *23*(5), 553–583. https://doi.org/10.1002/jae.1018

McCollister, K. E., French, M. T., & Fang, H. (2010). The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug and Alcohol Dependence*, *108*(1–2), 98–109. https://doi.org/10.1016/j.drugalcdep.2009.12.002

Meyer, D. (2010). Evaluating the Severity of Hate-motivated Violence: Intersectional Differences among LGBT Hate Crime Victims. *Sociology*, *44*(5), 980–995. https://doi.org/10.1177/0038038510375737

Miller, T. R., Cohen, M. A., & Rossman, S. B. (1993). Victim Costs Of Violent Crime And Resulting Injuries. *Health Affairs*, *12*(4), 186–197. https://doi.org/10.1377/hlthaff.12.4.186

Mills, C. E. (2019). Gay Visibility and Disorganized and Strained Communities: A Community-Level Analysis of Anti-Gay Hate Crime in New York City. *Journal of Interpersonal Violence*. Published. https://doi.org/10.1177/0886260519848784

Mills, C. E. (2020). A Common Target: Anti-Jewish Hate Crime in New York City Communities, 1995–2010. *Journal of Research in Crime and Delinquency*, *57*(6), 643–692. https://doi.org/10.1177/0022427820902832

Mindiola, T., Jr, Niemann, Y. F., Rodriguez, N., & Rodríguez, N. (2009). *Black-Brown Relations and Stereotypes*. Amsterdam University Press.

Morsy, S., Dang, T., Kamel, M., Zayan, A., Makram, O., Elhady, M., Hirayama, K., & Huy, N. (2018). Prediction of Zika-confirmed cases in Brazil and Colombia using Google Trends. *Epidemiology and Infection*, *146*(13), 1625–1627. https://doi.org/10.1017/s0950268818002078

Müller, K., & Schwarz, C. (2018). Making America Hate Again? Twitter and Hate Crime Under Trump. *SSRN Electronic Journal*. Published. https://doi.org/10.2139/ssrn.3149103

Müller, K., & Schwarz, C. (2020). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association*. Published. https://doi.org/10.1093/jeea/jvaa045

National Archives. (2022, January 24). *Executive Order 9066: Resulting in Japanese-American Incarceration*. Retrieved April 10, 2022, from https://www.archives.gov/milestone-documents/executive-order-9066

Newman, B., Merolla, J. L., Shah, S., Lemi, D. C., Collingwood, L., & Ramakrishnan, S. K. (2020). The Trump Effect: An Experimental Investigation of the Emboldening Effect of Racially Inflammatory Elite Communication. *British Journal of Political Science*, *51*(3), 1138–1159. https://doi.org/10.1017/s0007123419000590

Nitta, G. R., Rao, B. Y., Sravani, T., Ramakrishiah, N., & BalaAnand, M. (2019). LASSO-based feature selection and naïve Bayes classifier for crime prediction and its type. *Service Oriented Computing and Applications*, *13*(3), 187–197. https://doi.org/10.1007/s11761-018-0251-3

Nyamnjoh, F. B. (2010). Racism, Ethnicity and the Media in Africa: Reflections Inspired by Studies of Xenophobia in Cameroon and South Africa. *Africa Spectrum*, *45*(1), 57–93. https://doi.org/10.1177/000203971004500103

Perry, B. (2009). 'There's just places ya' don't wanna go': the segregating impact of hate crime against Native Americans. *Contemporary Justice Review*, *12*(4), 401–418. https://doi.org/10.1080/10282580903342888

Petropoulos, A., Siakoulis, V., Stavroulakis, E., Lazaris, P., & Vlachogiannakis, N. (2021). Employing Google Trends and Deep Learning in Forecasting Financial Market Turbulence. *Journal of Behavioral Finance*, 1–13. https://doi.org/10.1080/15427560.2021.1913160

Piatkowska, S. J., Messner, S. F., & Yang, T. C. (2018). Understanding the Relationship Between Relative Group Size and Hate Crime Rates: Linking Methods with Concepts. *Justice Quarterly*, *36*(6), 1072–1095. https://doi.org/10.1080/07418825.2018.1482942

Piña-García, C. A., & Ramírez-Ramírez, L. (2019). Exploring crime patterns in Mexico City. *Journal of Big Data*, *6*(65). https://doi.org/10.1186/s40537-019-0228-x

Powers, R. A., & Socia, K. M. (2018). Racial Animosity, Adversary Effect, and Hate Crime: Parsing Out Injuries in Intraracial, Interracial, and Race-Based Offenses. *Crime & Delinquency*, *65*(4), 447–473. https://doi.org/10.1177/0011128718779566

Prado-Román, C., Gómez-Martínez, R., & Orden-Cruz, C. (2020). Google Trends as a Predictor of Presidential Elections: The United States Versus Canada. *American Behavioral Scientist*, *65*(4), 666–680. https://doi.org/10.1177/0002764220975067

Shively, M. (2005). Study of literature and legislation on hate crime in America. Final report submitted to the National Institute of Justice, Washington, DC (NCJ 210300). Retrieved May 10, 2021 from https://nij.ojp.gov/library/publications/study-literature-and-legislation-hate-crime-america-final-report

Stacey, M., Carbone-López, K., & Rosenfeld, R. (2011). Demographic Change and Ethnically Motivated Crime: The Impact of Immigration on Anti-Hispanic Hate Crime in the United States. *Journal of Contemporary Criminal Justice*, *27*(3), 278–298. https://doi.org/10.1177/1043986211412560

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, *118*, 26–40. https://doi.org/10.1016/j.jpubeco.2014.04.010

Stop AAPI Hate. (2022, April). *Stop AAPI Hate National Report (Through December 31, 2021)*. https://stopaapihate.org/wp-content/uploads/2022/03/22-SAH-NationalReport-3.1.22-v9.pdf

Reier Forradellas, R. F., Náñez Alonso, S. L., Jorge-Vazquez, J., & Rodriguez, M. L. (2020). Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction. *Social Sciences*, *10*(1), 4. https://doi.org/10.3390/socsci10010004

Reilly, K. (2016, August 31). *Here Are All the Times Donald Trump Insulted Mexico*. Time. https://time.com/4473972/donald-trump-mexico-meeting-insult/

Rushin, S., & Edwards, G. S. (2018). The Effect of President Trump's Election on Hate Crimes. *SSRN Electronic Journal*. Published. https://doi.org/10.2139/ssrn.3102652

Tessler, H., Choi, M., & Kao, G. (2020). The Anxiety of Being Asian American: Hate Crimes and Negative Biases During the COVID-19 Pandemic. *American Journal of Criminal Justice*, *45*(4), 636–646. https://doi.org/10.1007/s12103-020-09541-5

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

United Nations Secretary-General. (2021, March 22). *Statement attributable to the Spokesperson for the Secretary-General*. https://www.un.org/sg/en/content/statement-attributable-the-spokesperson-for-the-secretary-general-the-rise-of-violence-against-asians

US Census Bureau. (2022, April 4). *Population*. Census.Gov. Retrieved April 29, 2022, from https://www.census.gov/topics/population.html

U.S. Congress. (1860) *U.S. Statutes at Large, Volume 12 -1861, 37th Congress*. United States, - 1861. Session 2, Chapter 27. [Periodical] Section Retrieved from the Library of Congress, https://www.loc.gov/item/llsl-v12/.

U.S. Congress. (1873) *U.S. Statutes at Large, Volume 18 -1875: 43rd Congress; Revised Statues in Force, Relating to D.C., and Post Roads; Public Treaties in Force*. United States, - 1875. Session 2, Chapter 141. [Periodical] Retrieved from the Library of Congress, https://www.loc.gov/item/llsl-v18/.

U.S. Congress. (1881) *U.S. Statutes at Large, Volume 22 -1883, 47th Congress*. United States, - 1883. Session 1 , Chapter 126. [Periodical] Retrieved from the Library of Congress, https://www.loc.gov/item/llsl-v22/.

U.S. Congress. (1892) *U.S. Statutes at Large, Volume 27 -1893, 52nd Congress*. United States, - 1893. Session 1, Chapter 60. [Periodical] Retrieved from the Library of Congress, https://www.loc.gov/item/llsl-v27/.

U.S. Congress. (1902) *U.S. Statutes at Large, Volume 32 -1903, 57th Congress*. United States, - 1903. Session 1, Chapter 641. [Periodical] Retrieved from the Library of Congress, https://www.loc.gov/item/llsl-v33/.

U.S. Congress. (1943) *U.S. Statutes at Large, Volume 57 , 78th Congress, Session 1*. United States. Session 1, Chapter 344. [Periodical] Retrieved from the Library of Congress, https://www.loc.gov/item/llsl-v57/.

U.S. Department of Agriculture (USDA). (2021, February 24). *USDA ERS - Download Data*. USDA. https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/

van Greunen, J., Heymans, A., van Heerden, C., & van Vuuren, G. (2014). The Prominence of Stationarity in Time Series Forecasting. *Studies in Economics and Econometrics*, *38*(1), 1–16. https://doi.org/10.1080/10800379.2014.12097260

Venkatesh, U., & Gandhi, P. A. (2020). Prediction of COVID-19 Outbreaks Using Google Trends in India: A Retrospective Analysis. *Healthcare Informatics Research*, *26*(3), 175–184. https://doi.org/10.4258/hir.2020.26.3.175

Vougas, D. V. (2008). On the size of the DF–GLS test. *Applied Economics Letters*, *15*(10), 777–780. https://doi.org/10.1080/13504850600749115

Wagner, U., Tachtsoglou, S., Kotzur, P. F., Friehs, M. T., & Kemmesies, U. (2020). Proportion of Foreigners Negatively Predicts the Prevalence of Xenophobic Hate Crimes within German Districts. *Social Psychology Quarterly*, *83*(2), 195–205. https://doi.org/10.1177/0190272519887719

Wang, J., Hu, J., Shen, S., Zhuang, J., & Ni, S. (2020). Crime risk analysis through big data algorithm with urban metrics. *Physica A: Statistical Mechanics and Its Applications*, *545*. https://doi.org/10.1016/j.physa.2019.123627

Welsh, B. C., Farrington, D. P., & Gowar, B. R. (2015). Benefit-Cost Analysis of Crime Prevention Programs. *Crime and Justice*, *44*(1), 447–516. https://doi.org/10.1086/681556

Westerlund, J. (2013). On the asymptotic distribution of the Dickey Fuller-GLS test statistic. *Statistics*, *48*(6), 1233–1253. https://doi.org/10.1080/02331888.2013.835815

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2019). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology*. Published. https://doi.org/10.1093/bjc/azz049

World Health Organization (WHO). (2020, January 30). *Pneumonia of unknown cause – China*. https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/

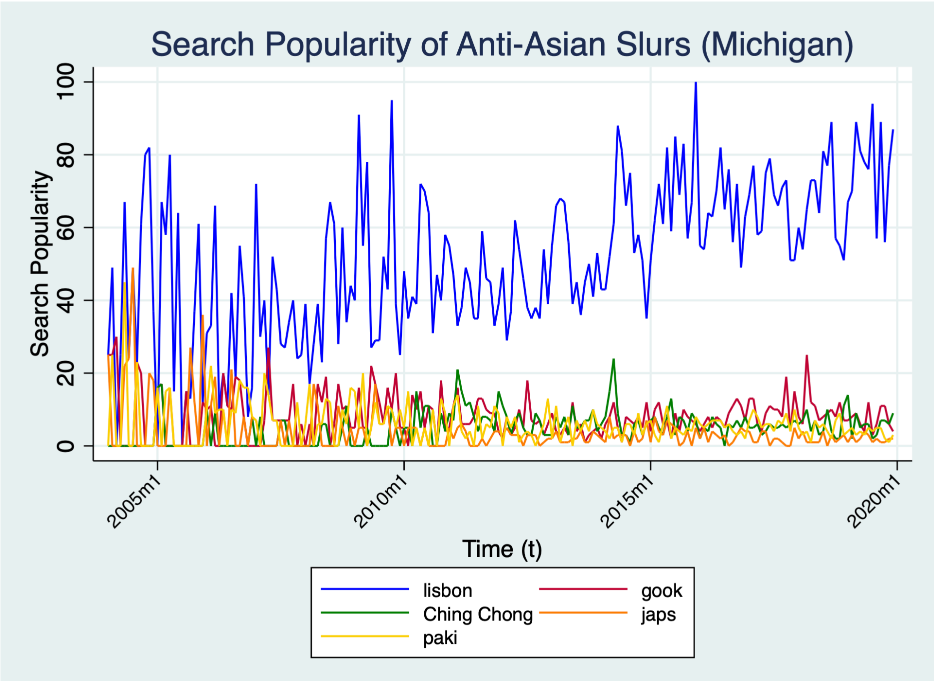
World Health Organization (WHO). (2022, April 28). *WHO Coronavirus (COVID-19) Dashboard*. WHO. Retrieved April 29, 2022, from https://covid19.who.int/

Wu, C., Qian, Y., & Wilkes, R. (2020). Anti-Asian discrimination and the Asian-white mental health gap during COVID-19. *Ethnic and Racial Studies*, *44*(5), 819–835. https://doi.org/10.1080/01419870.2020.1851739

Zhang, Y., Zhang, L., & Benton, F. (2021). Hate Crimes against Asian Americans. *American Journal of Criminal Justice*, 1–21. https://doi.org/10.1007/s12103-020-09602-9

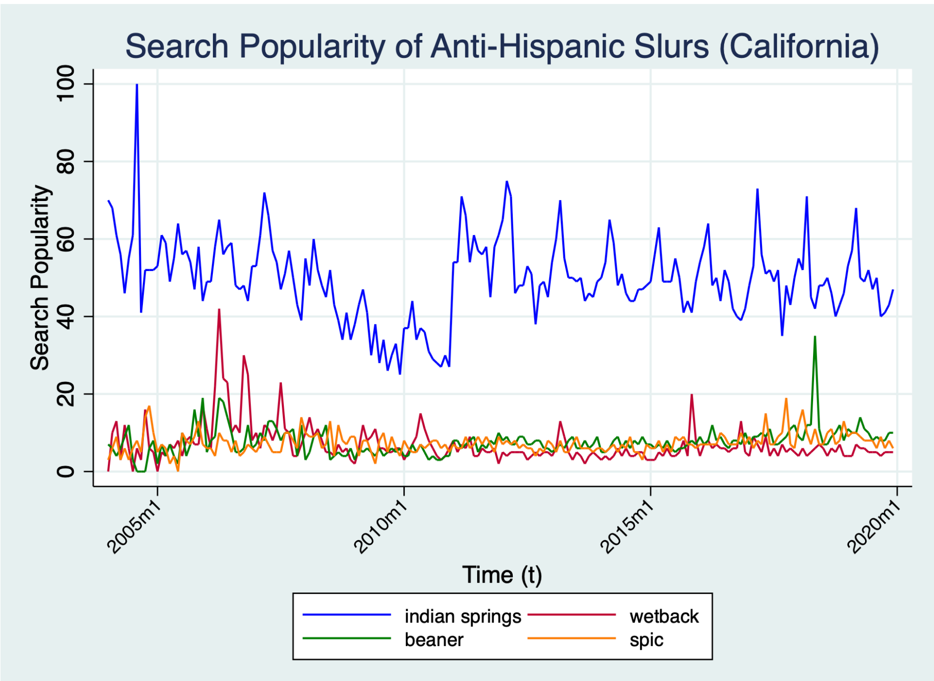
APPENDIX

Figure A.1: Search Popularity of Anti-Asian Slurs Including an Anchor Term (Michigan, 20014-2019)



Data Source: Constructed using data from Google Trends (https://www.google.com/trends)

Figure A.2: Search Popularity of Anti-Hispanic Slurs Including an Anchor Term (California, 20014-2019)



Data Source: Constructed using data from Google Trends (https://www.google.com/trends)

Table A.1: All Types of Crime Reported by the FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program

Crime	
Aggravated Assault	Motor Vehicle Theft
All Other Larceny	Murder and Nonnegligent Manslaughter
Animal Cruelty*	Negligent Manslaughter
Arson	Not Specified
Assisting or Promoting Prostitution*	Pocket-picking
Betting/Wagering*	Pornography/Obscene Material*
Bribery*	Prostitution
Burglary/Breaking & Entering	Purchasing Prostitution
Counterfeiting/Forgery	Purse-snatching
Credit Card/Automated Teller Machine Fraud	Rape
Destruction/Damage/Vandalism of Property	Robbery
Drug Equipment Violations	Sexual Assault With An Object
Drug/Narcotic Violations	Shoplifting
Embezzlement	Simple Assault
Extortion/Blackmail	Sodomy
False Pretences/Swindle/Confidence Game	Statutory Rape*
Fondling	Stolen Property Offenses
Hacking/Computer Invasion*	Theft From Building
Human Trafficking, Commercial Sex Acts*	Theft From Coin-Operated Machine or Device*
Identity Theft*	Theft From Motor Vehicle
Impersonation	Theft of Motor Vehicle Parts or Accessories
Incest*	Weapon Law Violations
Intimidation	Welfare Fraud*
Kidnapping/Abduction	Wire Fraud

*None of these crimes were committed with the motivational bias against Asians or Hispanics.

Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

Table A.2: All Types of Offender Bias Reported by the FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program

Offender Bias	
Anti-American Indian or Alaska Native	Anti-Lesbian (Female)
Anti-Arab	Anti-Lesbian, Gay, Bisexual, or Transgender (Mixed Group)
Anti-Asian	Anti-Male
Anti-Atheism/Agnosticism	Anti-Mental Disability
Anti-Bisexual	Anti-Mormon
Anti-Black or African American	Anti-Multiple Races, Group
Anti-Buddhist	Anti-Multiple Religions, Group
Anti-Catholic	Anti-Native Hawaiian or Other Pacific Islander
Anti-Eastern Orthodox (Russian, Greek, Other)	Anti-Other Christian
Anti-Female	Anti-Other Race/Ethnicity/Ancestry
Anti-Gay (Male)	Anti-Other Religion
Anti-Gender Non-Conforming	Anti-Physical Disability
Anti-Heterosexual	Anti-Protestant
Anti-Hindu	Anti-Sikh
Anti-Hispanic or Latino	Anti-Transgender
Anti-Islamic (Muslim)	Anti-White
Anti-Jehovah's Witness	Unknown (offender's motivation not known)
Anti-Jewish	

Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

Table A.3: All Locations of Crimes Reported by the FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program

Crime	
Field/Woods	Liquor Store
Residence/Home	Rental Storage Facility
Other/Unknown	Park/Playground
Highway/Road/Alley/Street/Sidewalk	School-College/University
Restaurant	School-Elementary/Secondary
School/College	Industrial Site
Bank/Savings and Loan	Shopping Mall
Church/Synagogue/Temple/Mosque	Gambling Facility/Casino/Race Track
Convenience Store	Camp/Campground
Lake/Waterway/Beach	Tribal Lands
Parking/Drop Lot/Garage	Arena/Stadium/Fairgrounds/Coliseum
Specialty Store	ATM Separate from Bank
Hotel/Motel/Etc.	Farm Facility
Service/Gas Station	Amusement Park
Bar/Nightclub	Daycare Facility
Jail/Prison/Penitentiary/Corrections Facility	Auto Dealership New/Used
Construction Site	Rest Area
Air/Bus/Train Terminal	Abandoned/Condemned Structure
Commercial/Office Building	Shelter-Mission/Homeless
Government/Public Building	Dock/Wharf/Freight/Modal Terminal
Grocery/Supermarket	Community Center
Drug Store/Doctor's Office/Hospital	Military Installation
Department/Discount Store	Cyberspace

Data Source: FBI's Uniform Crime Reporting (UCR) Hate Crime Statistics Program (FBI, n.d.)

<u>Table A.4: Correlation Matrix of All Variables Included in the Trump Extension (Section 6.2)</u>

	racist	racist^2	racists	racists^2	mexican	mexican^2	mexicans	mexicans^2	immigrant	immigrant^2	immigrants	immigrants^2	mexican wall	mexican wall^2	wetback	wetback^2	wetbacks	wetbacks^2	beaner	beaner^2	beaners	beaners^2	chopa	chopa^2	chopas	chopas^2	chopita	chopita^2	pepper belly	pepper belly^2	pepper bellys	pepper bellys^2	taco bender	taco bender^2	spic	spic^2	spics	spics^2	Total Employment	Employment (%)	Total Labour Force Participation	Labour Force Participation (%)	Total Unemployment	Unemployment (%)	Total Asian Population	Asian Population (%)	Total Hispanic Population	Hispanic Population (%)	Total Population	Anti-Hispanic Hate Crime (Lag 1)	Anti-Hispanic Hate Crime (Lag 2)	Anti-Hispanic Hate Crime (Lag 3)	Anti-Hispanic Hate Crime (Lag 4)	Anti-Hispanic Hate Crime (Lag 5)	Anti-Hispanic Hate Crime (Lag 6)
racist	1.00	0.93	0.39	0.39	-0.10	-0.13	-0.01	0.24	0.03	-0.03	0.04	0.04	-0.01	0.10	0.03	-0.01	0.00	-0.14	-0.09	-0.08	-0.03	-0.06	-0.04	-0.08	-0.04	0.02	0.02	-0.01	-0.01	-0.12	-0.04	-0.03	0.19	0.07	-0.07	0.07	-0.02	-0.02	0.17	0.00	0.03	0.00	-0.20	0.02	-0.38	-0.07	-0.48	-0.47	-0.36	0.06	0.00	0.04	0.01	0.01	0.03
racist^2	0.93	1.00	0.36	0.37	-0.18	-0.22	-0.03	0.19	0.01	-0.04	0.02	0.02	-0.01	0.06	0.04	0.00	0.00	-0.09	-0.06	-0.05	-0.03	-0.03	-0.02	-0.06	-0.03	0.01	0.03	-0.02	-0.04	-0.08	-0.05	-0.03	0.09	0.02	-0.03	0.04	0.00	-0.01	0.10	-0.01	0.05	-0.01	-0.10	0.01	-0.23	-0.03	-0.29	-0.28	-0.20	0.07	-0.01	0.05	-0.01	0.03	0.00
racists	0.39	0.36	1.00	0.95	-0.03	-0.02	-0.03	0.41	-0.08	0.12	0.01	0.02	-0.03	0.18	0.08	0.04	0.04	0.04	-0.01	0.00	0.01	0.07	-0.06	-0.05	0.13	0.13	0.07	0.05	0.05	0.02	-0.12	-0.08	0.09	0.00	-0.09	0.10	-0.01	-0.05	0.16	-0.08	0.11	-0.07	-0.14	-0.04	-0.12	0.14	-0.19	-0.14	-0.26	0.06	-0.02	-0.01	-0.09	-0.02	0.01
racists^2	0.39	0.37	0.95	1.00	0.00	0.00	0.05	0.43	-0.07	0.10	0.05	0.05	-0.02	0.16	0.07	0.04	0.05	0.08	-0.04	-0.03	0.00	0.05	-0.10	-0.06	0.10	0.09	0.06	0.04	-0.01	-0.03	-0.10	-0.07	0.06	-0.01	-0.06	0.11	-0.01	-0.04	0.14	-0.02	0.09	-0.05	-0.12	-0.06	-0.06	0.15	-0.12	-0.07	-0.21	0.03	-0.02	0.01	-0.07	-0.04	0.05
mexican	-0.10	-0.18	-0.03	0.00	1.00	0.98	0.44	0.06	0.05	-0.07	0.16	0.09	0.04	0.00	-0.08	-0.02	0.17	0.09	0.01	0.03	-0.05	-0.14	-0.06	-0.03	-0.06	0.08	-0.08	-0.11	0.05	0.05	-0.02	-0.04	0.05	0.03	0.06	0.05	-0.06	-0.04	0.00	0.03	0.01	0.08	0.01	0.06	0.00	-0.01	0.00	-0.01	0.02	-0.30	0.16	-0.16	0.11	-0.04	0.07
mexican^2	-0.13	-0.22	-0.02	0.00	0.98	1.00	0.38	0.06	0.01	-0.08	0.12	0.06	0.05	0.00	-0.06	-0.02	0.17	0.08	0.03	0.04	-0.05	-0.12	-0.03	-0.01	-0.03	0.09	-0.07	-0.10	0.03	0.03	-0.02	-0.03	0.05	0.03	0.05	0.05	-0.05	-0.04	0.00	0.03	0.00	0.07	0.01	0.04	0.00	0.00	0.00	0.00	0.01	-0.30	0.18	-0.17	0.10	-0.05	0.07
mexicans	-0.01	-0.03	-0.03	0.05	0.44	0.38	1.00	0.22	0.31	-0.02	0.54	0.42	0.27	0.27	-0.11	-0.12	0.04	0.05	0.05	0.06	0.03	-0.05	-0.21	-0.16	-0.06	-0.02	-0.07	-0.02	-0.01	-0.02	-0.15	-0.17	0.01	-0.02	0.05	-0.04	-0.05	0.03	0.00	0.08	0.01	0.12	0.02	0.07	0.01	0.00	0.01	0.00	0.01	-0.17	0.02	-0.08	0.16	-0.10	0.10
mexicans^2	0.24	0.19	0.41	0.43	0.06	0.06	0.22	1.00	0.01	0.32	0.08	0.10	-0.04	0.24	-0.10	0.01	-0.08	-0.06	0.07	0.06	0.05	0.03	-0.08	-0.12	0.11	0.05	-0.01	0.01	-0.01	-0.06	-0.13	-0.13	0.12	0.09	-0.17	0.00	0.00	0.04	0.26	0.09	0.07	0.06	-0.28	-0.05	-0.05	0.25	-0.24	-0.20	-0.26	0.11	0.08	-0.10	0.01	-0.13	-0.08
immigrant	0.03	0.01	-0.08	-0.07	0.05	0.01	0.31	0.01	1.00	0.16	0.76	0.66	0.14	0.17	0.00	0.03	0.03	0.00	-0.08	-0.07	-0.13	-0.18	-0.03	0.01	-0.05	0.07	-0.03	-0.01	-0.02	-0.02	0.09	0.10	-0.01	-0.05	-0.03	-0.04	0.14	0.15	0.00	-0.01	0.00	-0.02	-0.01	-0.05	0.00	-0.02	0.00	-0.02	0.03	0.01	-0.14	0.07	0.06	-0.06	0.03
immigrant^2	-0.03	-0.04	0.12	0.10	-0.07	-0.08	-0.02	0.32	0.16	1.00	0.07	0.08	-0.11	0.17	-0.13	0.00	-0.03	0.16	0.24	0.22	-0.05	-0.05	-0.02	0.00	0.12	0.08	-0.03	-0.09	-0.09	-0.04	0.00	0.04	0.09	0.06	0.04	0.02	0.09	0.10	0.11	-0.02	0.20	0.00	-0.04	0.03	-0.02	-0.13	0.17	0.21	-0.01	0.18	0.02	-0.02	0.00	-0.02	-0.10
immigrants	0.04	0.02	0.01	0.05	0.16	0.12	0.54	0.08	0.76	0.07	1.00	0.92	0.15	0.27	0.06	0.10	0.08	0.00	-0.06	-0.06	-0.15	-0.15	-0.13	-0.06	-0.10	0.00	0.01	0.02	0.01	0.00	0.00	0.01	0.04	0.00	0.01	0.00	0.10	0.12	-0.01	-0.03	0.00	0.00	0.01	-0.02	0.00	-0.02	-0.01	-0.02	0.04	-0.10	-0.09	0.14	0.01	-0.06	0.13
immigrants^2	0.04	0.02	0.02	0.05	0.09	0.06	0.42	0.10	0.66	0.08	0.92	1.00	-0.03	0.21	0.08	0.11	0.10	0.03	-0.03	-0.03	-0.16	-0.15	-0.09	-0.03	-0.08	0.01	0.00	0.01	0.02	0.01	-0.02	-0.01	0.08	0.05	0.01	0.01	0.06	0.07	0.00	-0.01	0.00	0.01	0.01	-0.03	0.00	-0.01	0.00	-0.01	0.02	-0.06	-0.09	0.17	-0.06	-0.03	0.06
mexican wall	-0.01	-0.01	-0.03	-0.02	0.04	0.05	0.27	-0.04	0.14	-0.11	0.15	-0.03	1.00	0.38	-0.03	-0.02	-0.15	-0.06	-0.08	-0.05	0.07	0.03	-0.07	-0.05	0.07	0.10	-0.04	0.00	-0.01	0.01	0.05	0.08	-0.10	-0.15	0.07	0.00	0.10	0.06	-0.01	-0.02	-0.01	-0.05	0.01	-0.04	0.00	-0.02	0.00	-0.02	0.04	-0.10	0.10	-0.08	0.06	-0.08	0.17
mexican wall^2	0.10	0.06	0.18	0.16	0.00	0.00	0.27	0.24	0.17	0.17	0.27	0.21	0.38	1.00	0.01	-0.01	-0.04	-0.07	-0.03	-0.02	0.01	0.01	-0.02	-0.02	-0.03	0.00	-0.01	-0.05	-0.04	-0.05	-0.05	-0.03	0.17	0.08	-0.05	0.01	0.00	0.00	0.11	0.05	0.08	0.05	-0.12	-0.06	-0.17	-0.03	-0.14	-0.10	-0.19	-0.08	0.06	0.00	0.01	-0.13	0.09
wetback	0.03	0.04	0.08	0.07	-0.08	-0.06	-0.11	-0.10	0.00	-0.13	0.06	0.08	-0.03	0.01	1.00	0.64	0.07	0.04	-0.11	-0.11	0.12	0.17	0.13	0.09	-0.09	-0.01	-0.07	-0.06	-0.03	-0.06	0.04	0.06	0.06	0.08	0.00	0.12	-0.09	-0.14	-0.01	-0.09	-0.01	-0.09	-0.01	-0.10	0.00	0.01	0.00	-0.01	0.00	-0.04	-0.07	0.11	-0.18	0.06	0.19
wetback^2	-0.01	0.00	0.04	0.04	-0.02	-0.02	-0.12	0.01	0.03	0.00	0.10	0.11	-0.02	-0.01	0.64	1.00	0.13	0.13	-0.02	-0.01	0.03	0.07	-0.01	0.04	-0.20	-0.11	0.00	-0.01	0.13	0.11	0.01	0.01	0.01	0.02	-0.04	0.04	0.01	-0.07	0.01	-0.02	0.00	0.02	-0.02	0.00	0.01	0.01	0.00	0.01	-0.01	-0.08	-0.08	0.07	-0.12	-0.13	0.17
wetbacks	0.00	0.00	0.04	0.05	0.17	0.17	0.04	-0.08	0.03	-0.03	0.08	0.10	-0.15	-0.04	0.07	0.13	1.00	0.64	0.03	0.02	-0.15	-0.13	-0.23	-0.08	-0.19	-0.07	-0.05	-0.04	0.04	0.00	-0.19	-0.18	-0.03	-0.02	0.14	0.31	-0.04	-0.04	-0.01	-0.03	-0.01	0.02	0.01	0.04	0.01	0.00	0.01	0.00	0.02	-0.19	-0.04	0.07	-0.03	0.08	-0.03
wetbacks^2	-0.14	-0.09	0.04	0.08	0.09	0.08	0.05	-0.06	0.00	0.16	0.00	0.03	-0.06	-0.07	0.04	0.13	0.64	1.00	0.04	0.04	-0.10	-0.10	-0.23	-0.10	-0.02	0.02	0.00	-0.04	-0.01	0.07	-0.08	-0.06	-0.16	-0.11	0.32	0.24	0.01	-0.01	0.03	0.00	0.02	0.02	-0.02	0.02	0.30	0.09	0.34	0.31	0.25	-0.04	-0.07	0.01	-0.02	0.02	0.03
beaner	-0.09	-0.06	-0.01	-0.04	0.01	0.03	0.05	0.07	-0.08	0.24	-0.06	-0.03	-0.08	-0.03	-0.11	-0.02	0.03	0.04	1.00	0.95	0.01	0.03	-0.06	-0.07	0.08	-0.05	-0.04	-0.10	-0.06	0.01	-0.07	-0.07	-0.14	-0.12	0.01	0.00	-0.16	-0.12	0.01	-0.11	0.11	-0.09	0.04	0.00	0.08	-0.09	0.25	0.27	0.09	0.00	0.06	0.01	-0.03	-0.12	-0.07
beaner^2	-0.08	-0.05	0.00	-0.03	0.03	0.04	0.06	0.06	-0.07	0.22	-0.06	-0.03	-0.05	-0.02	-0.11	-0.01	0.02	0.04	0.95	1.00	-0.05	-0.03	-0.04	-0.06	0.07	-0.08	-0.05	-0.08	-0.06	0.00	-0.06	-0.05	-0.14	-0.11	0.03	0.03	-0.13	-0.08	0.04	-0.07	0.12	-0.07	0.00	-0.02	0.09	-0.08	0.24	0.26	0.10	0.01	0.08	0.00	-0.01	-0.11	-0.07
beaners	-0.03	-0.03	0.01	0.00	-0.05	-0.05	0.03	0.05	-0.13	-0.05	-0.15	-0.16	0.07	0.01	0.12	0.03	-0.15	-0.10	0.01	-0.05	1.00	0.91	-0.01	-0.01	0.13	0.15	-0.01	0.02	0.04	0.06	0.16	0.16	-0.01	0.00	-0.04	-0.19	-0.14	-0.14	0.00	-0.03	-0.01	-0.05	-0.01	-0.01	0.00	0.00	0.00	0.00	0.01	-0.15	0.13	-0.19	0.17	-0.12	0.11
beaners^2	-0.06	-0.03	0.07	0.05	-0.14	-0.12	-0.05	0.03	-0.18	-0.05	-0.15	-0.15	0.03	0.01	0.17	0.07	-0.13	-0.10	0.03	-0.03	0.91	1.00	0.02	0.00	0.14	0.11	-0.01	0.01	0.02	0.04	0.13	0.12	-0.01	0.00	-0.07	-0.16	-0.09	-0.07	0.00	-0.05	0.00	-0.06	-0.01	-0.01	0.00	-0.01	0.00	0.00	0.00	-0.16	0.15	-0.20	0.14	-0.07	0.09
chopa	-0.04	-0.02	-0.06	-0.10	-0.06	-0.03	-0.21	-0.08	-0.03	-0.02	-0.13	-0.09	-0.07	-0.02	0.13	-0.01	-0.23	-0.23	-0.06	-0.04	-0.01	0.02	1.00	0.70	-0.06	0.02	-0.14	-0.06	-0.01	-0.03	0.22	0.25	0.03	0.01	-0.26	-0.39	0.05	0.00	0.02	-0.01	0.02	0.02	-0.01	-0.03	0.01	0.00	0.01	0.01	0.01	0.03	0.04	0.00	-0.07	0.03	-0.03
chopa^2	-0.08	-0.06	-0.05	-0.06	-0.03	-0.01	-0.16	-0.12	0.01	0.00	-0.06	-0.03	-0.05	-0.02	0.09	0.04	-0.08	-0.10	-0.07	-0.06	-0.01	0.00	0.70	1.00	-0.05	0.04	-0.08	-0.07	-0.09	-0.07	0.39	0.40	-0.01	-0.01	-0.25	-0.24	0.09	0.08	0.01	-0.03	-0.01	-0.02	-0.03	-0.08	0.06	-0.04	0.15	0.15	0.07	0.02	0.00	0.04	-0.01	-0.07	-0.09
chopas	-0.04	-0.03	0.13	0.10	-0.06	-0.03	-0.06	0.11	-0.05	0.12	-0.10	-0.08	0.07	-0.03	-0.09	-0.20	-0.19	-0.02	0.08	0.07	0.13	0.14	-0.06	-0.05	1.00	0.67	0.07	0.07	-0.06	-0.06	-0.04	-0.05	-0.06	-0.06	-0.13	-0.03	-0.05	-0.10	0.06	-0.04	0.04	-0.16	-0.05	-0.23	0.03	-0.05	0.09	0.08	0.07	0.20	-0.06	0.04	0.06	-0.01	-0.04
chopas^2	0.02	0.01	0.13	0.09	0.08	0.09	-0.02	0.05	0.07	0.08	0.00	0.01	0.10	0.00	-0.01	-0.11	-0.07	0.02	-0.05	-0.08	0.15	0.11	0.02	0.04	0.67	1.00	0.03	0.02	0.03	0.02	0.00	0.00	-0.01	-0.03	-0.11	-0.09	-0.04	-0.18	0.00	0.01	-0.01	-0.07	0.00	-0.18	0.01	-0.01	0.00	0.00	0.02	0.10	-0.18	0.10	-0.04	-0.01	0.01
chopita	0.02	0.03	0.07	0.06	-0.08	-0.07	-0.07	-0.01	-0.03	-0.03	0.01	0.00	-0.04	-0.01	-0.07	0.00	-0.05	0.00	-0.04	-0.05	-0.01	-0.01	-0.14	-0.08	0.07	0.03	1.00	0.52	0.08	0.05	0.07	0.02	-0.05	-0.02	0.01	0.03	0.01	0.01	0.01	0.01	0.01	-0.02	0.00	-0.04	-0.01	-0.03	0.02	0.02	0.00	0.07	-0.02	0.10	-0.08	-0.11	0.13
chopita^2	-0.01	-0.02	0.05	0.04	-0.11	-0.10	-0.02	0.01	-0.01	-0.09	0.02	0.01	0.00	-0.05	-0.06	-0.01	-0.04	-0.04	-0.10	-0.08	0.02	0.01	-0.06	-0.07	0.07	0.02	0.52	1.00	0.06	0.02	0.04	-0.01	-0.09	-0.06	-0.08	0.02	0.02	0.02	0.00	-0.02	-0.17	-0.04	-0.10	0.00	0.01	0.08	-0.09	-0.09	-0.04	0.04	0.00	0.08	0.01	0.10	0.02
pepper belly	-0.01	-0.04	0.05	-0.01	0.05	0.03	-0.01	-0.01	-0.02	-0.09	0.01	0.02	-0.01	-0.04	-0.03	0.13	0.04	-0.01	-0.06	-0.06	0.04	0.02	-0.01	-0.09	-0.06	0.03	0.08	0.06	1.00	0.91	-0.11	-0.12	0.08	0.04	0.09	-0.06	-0.12	-0.19	-0.02	-0.05	-0.09	0.05	-0.03	0.16	-0.13	-0.03	-0.17	-0.16	-0.15	-0.06	-0.04	-0.06	-0.02	-0.04	0.02
pepper belly^2	-0.12	-0.08	0.02	-0.03	0.05	0.03	-0.02	-0.06	-0.02	-0.04	0.00	0.01	0.01	-0.05	-0.06	0.11	0.00	0.07	0.01	0.00	0.06	0.04	-0.03	-0.07	-0.06	0.02	0.05	0.02	0.91	1.00	-0.07	-0.07	-0.03	-0.01	0.22	-0.08	-0.13	-0.20	-0.02	-0.03	-0.02	0.08	0.02	0.19	0.07	0.02	0.08	0.07	0.03	-0.05	-0.01	-0.05	-0.03	-0.02	0.01
pepper bellys	-0.04	-0.05	-0.12	-0.10	-0.02	-0.02	-0.15	-0.13	0.09	0.00	0.00	-0.02	0.05	-0.05	0.04	0.01	-0.19	-0.08	-0.07	-0.06	0.16	0.13	0.22	0.39	-0.04	0.00	0.07	0.04	-0.11	-0.07	1.00	0.93	-0.14	-0.09	0.03	-0.13	0.27	0.27	0.03	-0.13	0.03	-0.15	-0.01	-0.07	0.16	0.03	0.18	0.16	0.14	0.04	-0.05	-0.07	0.09	-0.14	0.08
pepper bellys^2	-0.03	-0.03	-0.08	-0.07	-0.04	-0.03	-0.17	-0.13	0.10	0.04	0.01	-0.01	0.08	-0.03	0.06	0.01	-0.18	-0.06	-0.07	-0.05	0.16	0.12	0.25	0.40	-0.05	0.00	0.02	-0.01	-0.12	-0.07	0.93	1.00	-0.09	-0.05	0.07	-0.15	0.33	0.31	0.06	-0.06	0.10	-0.08	-0.02	-0.02	0.17	0.05	0.20	0.19	0.12	0.07	-0.08	-0.03	0.11	-0.12	0.04
taco bender	0.19	0.09	0.09	0.06	0.05	0.05	0.01	0.12	-0.01	0.09	0.04	0.08	-0.10	0.17	0.06	0.01	-0.03	-0.16	-0.14	-0.14	-0.01	-0.01	0.03	-0.01	-0.06	-0.01	-0.05	-0.09	0.08	-0.03	-0.14	-0.09	1.00	0.89	-0.06	0.05	0.01	0.01	-0.04	0.08	-0.04	0.01	-0.06	-0.17	-0.34	-0.11	-0.32	-0.26	-0.33	-0.08	0.03	0.13	-0.11	0.04	-0.03
taco bender^2	0.07	0.02	0.00	-0.01	0.03	0.03	-0.02	0.09	-0.05	0.06	0.00	0.05	-0.15	0.08	0.08	0.02	-0.02	-0.11	-0.12	-0.11	0.00	0.00	0.01	-0.01	-0.06	-0.03	-0.02	-0.06	0.04	-0.01	-0.09	-0.05	0.89	1.00	-0.01	0.06	0.00	0.00	-0.12	0.08	-0.02	0.00	0.05	-0.18	-0.18	-0.08	-0.13	-0.09	-0.15	-0.08	0.06	0.11	-0.10	0.01	0.01
spic	-0.07	-0.03	-0.09	-0.06	0.06	0.05	0.05	-0.17	-0.03	0.04	0.01	0.01	0.07	-0.05	0.00	-0.04	0.14	0.32	0.01	0.03	-0.04	-0.07	-0.26	-0.25	-0.13	-0.11	0.01	-0.08	0.09	0.22	0.03	0.07	-0.06	-0.01	1.00	0.41	0.06	0.03	-0.13	-0.04	0.04	0.02	0.19	0.18	0.14	-0.08	0.26	0.24	0.21	-0.09	0.01	0.00	0.01	0.04	0.04
spic^2	0.07	0.04	0.10	0.11	0.05	0.05	-0.04	0.00	-0.04	0.02	0.00	0.01	0.00	0.01	0.12	0.04	0.31	0.24	0.00	0.03	-0.19	-0.16	-0.39	-0.24	-0.03	-0.09	0.03	0.02	-0.06	-0.08	-0.13	-0.15	0.05	0.06	0.41	1.00	0.03	0.06	0.00	-0.01	-0.01	-0.05	-0.02	-0.05	0.01	0.01	0.00	0.01	0.00	-0.13	0.10	-0.02	-0.01	0.01	-0.02
spics	-0.02	0.00	-0.01	-0.01	-0.06	-0.05	-0.05	0.00	0.14	0.09	0.10	0.06	0.10	0.00	-0.09	0.01	-0.04	0.01	-0.16	-0.13	-0.14	-0.09	0.05	0.09	-0.05	-0.04	0.01	0.02	-0.12	-0.13	0.27	0.33	0.01	0.00	0.06	0.03	1.00	0.92	0.01	0.02	-0.01	-0.05	-0.03	-0.09	0.02	0.02	0.01	0.01	-0.01	0.03	-0.15	0.13	-0.03	0.04	-0.19
spics^2	-0.02	-0.01	-0.05	-0.04	-0.04	-0.04	0.03	0.04	0.15	0.10	0.12	0.07	0.06	0.00	-0.14	-0.07	-0.04	-0.01	-0.12	-0.08	-0.14	-0.07	0.00	0.08	-0.10	-0.18	0.01	0.02	-0.19	-0.20	0.27	0.31	0.01	0.00	0.03	0.06	0.92	1.00	0.01	0.00	0.00	-0.06	-0.03	-0.08	0.01	0.02	0.01	0.01	-0.03	0.04	-0.09	0.04	0.02	0.03	-0.20
Total Employment	0.17	0.10	0.16	0.14	0.00	0.00	0.00	0.26	0.00	0.11	-0.01	0.00	-0.01	0.11	-0.01	0.01	-0.01	0.03	0.01	0.04	0.00	0.00	0.02	0.01	0.06	0.00	0.01	0.00	-0.02	-0.02	0.03	0.06	-0.04	-0.12	-0.13	0.00	0.01	0.01	1.00	0.06	0.56	0.03	-0.82	-0.06	-0.01	0.21	-0.24	-0.28	-0.19	0.02	0.02	0.03	0.00	0.02	0.04
Employment (%)	0.00	-0.01	-0.08	-0.02	0.03	0.03	0.08	0.09	-0.01	-0.02	-0.03	-0.01	-0.02	0.05	-0.09	-0.02	-0.03	0.00	-0.11	-0.07	-0.03	-0.05	-0.01	-0.03	-0.04	0.01	0.01	-0.02	-0.05	-0.03	-0.13	-0.06	0.08	0.08	-0.04	-0.01	0.02	0.00	0.06	1.00	0.06	0.82	-0.06	-0.09	-0.01	-0.01	-0.02	-0.02	0.00	-0.01	0.03	-0.02	0.04	0.01	-0.04
Total Labour Force Participation	0.03	0.05	0.11	0.09	0.01	0.00	0.01	0.07	0.00	0.20	0.00	0.00	-0.01	0.08	-0.01	0.00	-0.01	0.02	0.11	0.12	-0.01	0.00	0.02	-0.01	0.04	-0.01	0.01	-0.17	-0.09	-0.02	0.03	0.10	-0.04	-0.02	0.04	-0.01	-0.01	0.00	0.56	0.06	1.00	0.10	-0.01	0.07	0.04	-0.03	0.15	0.15	0.07	0.04	0.05	0.03	0.00	0.02	0.06
Labour Force Participation (%)	0.00	-0.01	-0.07	-0.05	0.08	0.07	0.12	0.06	-0.02	0.00	0.00	0.01	-0.05	0.05	-0.09	0.02	0.02	0.02	-0.09	-0.07	-0.05	-0.06	0.02	-0.02	-0.16	-0.07	-0.02	-0.04	0.05	0.08	-0.15	-0.08	0.01	0.00	0.02	-0.05	-0.05	-0.06	0.03	0.82	0.10	1.00	0.04	0.39	0.01	0.01	-0.01	-0.01	0.00	0.01	0.01	-0.04	0.03	0.07	-0.11
Total Unemployment	-0.20	-0.10	-0.14	-0.12	0.01	0.01	0.02	-0.28	-0.01	-0.04	0.01	0.01	0.01	-0.12	-0.01	-0.02	0.01	-0.02	0.04	0.00	-0.01	-0.01	-0.01	-0.03	-0.05	0.00	0.00	-0.10	-0.03	0.02	-0.01	-0.02	-0.06	0.05	0.19	-0.02	-0.03	-0.03	-0.82	-0.06	-0.01	0.04	1.00	0.18	0.09	-0.26	0.41	0.43	0.33	0.00	0.01	-0.02	0.02	0.01	0.00
Unemployment (%)	0.02	0.01	-0.04	-0.06	0.06	0.04	0.07	-0.05	-0.05	0.03	-0.02	-0.03	-0.04	-0.06	-0.10	0.00	0.04	0.02	0.00	-0.02	-0.01	-0.01	-0.03	-0.08	-0.23	-0.18	-0.04	0.00	0.16	0.19	-0.07	-0.02	-0.17	-0.18	0.18	-0.05	-0.09	-0.08	-0.06	-0.09	0.07	0.39	0.18	1.00	0.02	0.04	0.01	0.01	0.00	0.02	-0.01	-0.07	0.08	0.00	-0.06
Total Asian Population	-0.38	-0.23	-0.12	-0.06	0.00	0.00	0.01	-0.05	0.00	-0.02	0.00	0.00	0.00	-0.17	0.00	0.01	0.01	0.30	0.08	0.09	0.00	0.00	0.01	0.06	0.03	0.01	-0.01	0.01	-0.13	0.07	0.16	0.17	-0.34	-0.18	0.14	0.01	0.02	0.01	-0.01	-0.01	0.04	0.01	0.09	0.02	1.00	0.70	0.74	0.64	0.74	-0.01	-0.02	-0.03	-0.02	0.00	0.00
Asian Population (%)	-0.07	-0.03	0.14	0.15	-0.01	0.00	0.00	0.25	-0.02	-0.13	-0.02	-0.01	-0.02	-0.03	0.01	0.01	0.00	0.09	-0.09	-0.08	0.00	-0.01	0.00	-0.04	-0.05	-0.01	-0.03	0.08	-0.03	0.02	0.03	0.05	-0.11	-0.08	-0.08	0.01	0.02	0.02	0.21	-0.01	-0.03	0.01	-0.26	0.04	0.70	1.00	0.12	0.09	0.11	0.02	-0.02	-0.08	-0.07	-0.07	-0.07
Total Hispanic Population	-0.48	-0.29	-0.19	-0.12	0.00	0.00	0.01	-0.24	0.00	0.17	-0.01	0.00	0.00	-0.14	0.00	0.00	0.01	0.34	0.25	0.24	0.00	0.00	0.01	0.15	0.09	0.00	0.02	-0.09	-0.17	0.08	0.18	0.20	-0.32	-0.13	0.26	0.00	0.01	0.01	-0.24	-0.02	0.15	-0.01	0.41	0.01	0.74	0.12	1.00	0.97	0.79	0.00	0.00	-0.01	-0.01	0.00	0.00
Hispanic Population (%)	-0.47	-0.28	-0.14	-0.07	-0.01	0.00	0.00	-0.20	-0.02	0.21	-0.02	-0.01	-0.02	-0.10	-0.01	0.01	0.00	0.31	0.27	0.26	0.00	0.00	0.01	0.15	0.08	0.00	0.02	-0.09	-0.16	0.07	0.16	0.19	-0.26	-0.09	0.24	0.01	0.01	0.01	-0.28	-0.02	0.15	-0.01	0.43	0.01	0.64	0.09	0.97	1.00	0.62	0.02	0.00	-0.04	-0.05	-0.05	-0.05
Total Population	-0.36	-0.20	-0.26	-0.21	0.02	0.01	0.01	-0.26	0.03	-0.01	0.04	0.02	0.04	-0.19	0.00	-0.01	0.02	0.25	0.09	0.10	0.01	0.00	0.01	0.07	0.07	0.02	0.00	-0.04	-0.15	0.03	0.14	0.12	-0.33	-0.15	0.21	0.00	-0.01	-0.03	-0.19	0.00	0.07	0.00	0.33	0.00	0.74	0.11	0.79	0.62	1.00	-0.06	-0.01	0.07	0.09	0.10	0.12
Anti-Hispanic Hate Crime (Lag 1)	0.06	0.07	0.06	0.03	-0.30	-0.30	-0.17	0.11	0.01	0.18	-0.10	-0.06	-0.10	-0.08	-0.04	-0.08	-0.19	-0.04	0.00	0.01	-0.15	-0.16	0.03	0.02	0.20	0.10	0.07	0.04	-0.06	-0.05	0.04	0.07	-0.08	-0.08	-0.09	-0.13	0.03	0.04	0.02	-0.01	0.04	0.01	0.00	0.02	-0.01	0.02	0.00	0.02	-0.06	1.00	-0.38	0.04	-0.06	0.07	-0.16
Anti-Hispanic Hate Crime (Lag 2)	0.00	-0.01	-0.02	-0.02	0.16	0.18	0.02	0.08	-0.14	0.02	-0.09	-0.09	0.10	0.06	-0.07	-0.08	-0.04	-0.07	0.06	0.08	0.13	0.15	0.04	0.00	-0.06	-0.18	-0.02	0.00	-0.04	-0.01	-0.05	-0.08	0.03	0.06	0.01	0.10	-0.15	-0.09	0.02	0.03	0.05	0.01	0.01	-0.01	-0.02	-0.02	0.00	0.00	-0.01	-0.38	1.00	-0.38	0.04	-0.06	0.07
Anti-Hispanic Hate Crime (Lag 3)	0.04	0.05	-0.01	0.01	-0.16	-0.17	-0.08	-0.10	0.07	-0.02	0.14	0.17	-0.08	0.00	0.11	0.07	0.07	0.01	0.01	0.00	-0.19	-0.20	0.00	0.04	0.04	0.10	0.10	0.08	-0.06	-0.05	-0.07	-0.03	0.13	0.11	0.00	-0.02	0.13	0.04	0.03	-0.02	0.03	-0.04	-0.02	-0.07	-0.03	-0.08	-0.01	-0.04	0.07	0.04	-0.38	1.00	-0.38	0.04	-0.06
Anti-Hispanic Hate Crime (Lag 4)	0.01	-0.01	-0.09	-0.07	0.11	0.10	0.16	0.01	0.06	0.00	0.01	-0.06	0.06	0.01	-0.18	-0.12	-0.03	-0.02	-0.03	-0.01	0.17	0.14	-0.07	-0.01	0.06	-0.04	-0.08	0.01	-0.02	-0.03	0.09	0.11	-0.11	-0.10	0.01	-0.01	-0.03	0.02	0.00	0.04	0.00	0.03	0.02	0.08	-0.02	-0.07	-0.01	-0.05	0.09	-0.06	0.04	-0.38	1.00	-0.37	0.04
Anti-Hispanic Hate Crime (Lag 5)	0.01	0.03	-0.02	-0.04	-0.04	-0.05	-0.10	-0.13	-0.06	-0.02	-0.06	-0.03	-0.08	-0.13	0.06	-0.13	0.08	0.02	-0.12	-0.11	-0.12	-0.07	0.03	-0.07	-0.01	-0.01	-0.11	-0.10	-0.04	-0.02	-0.14	-0.12	0.04	0.01	0.04	0.01	0.04	0.03	0.02	0.01	0.02	0.07	0.01	0.00	0.00	-0.07	0.00	-0.05	0.10	0.07	-0.06	0.04	-0.37	1.00	-0.38
Anti-Hispanic Hate Crime (Lag 6)	0.03	0.00	0.01	0.05	0.07	0.07	0.10	-0.08	0.03	-0.10	0.13	0.06	0.17	0.09	0.19	0.17	-0.03	0.03	-0.07	-0.07	0.11	0.09	-0.03	-0.09	-0.04	0.01	0.13	0.02	0.02	0.01	0.08	0.04	-0.03	0.01	0.04	-0.02	-0.19	-0.20	0.04	-0.04	0.06	-0.11	0.00	-0.06	0.00	-0.07	0.00	-0.05	0.12	-0.16	0.07	-0.06	0.04	-0.38	1.00