**University of Amsterdam, AY 2022/2023**

**Mattia Guarnerio, Student Number: 14350920, e-mail:** [**mattia.guarnerio@student.uva.nl**](mailto:mattia.guarnerio@student.uva.nl)

**Final Report for *Big Data & Automated Content Analysis, Part I & II*; Research Master’s Social Sciences, Year 1, Semester 2; Dr. D. C. Trilling**

**White Noise**

*The Radicalization of the US Republican Party’s Legislative Activity (2008–2016)*

**Table of Contents**

Introduction, and Theoretical Framework, p. 2

Analytical Strategy: Data Collection, p. 4

Analytical Strategy: Supervised Machine Learning, p. 6

Exploratory Analysis, Statistical Analysis, and Results, p. 9

Discussion, Conclusions, and Limitations of the Study, p. 13

References, p. 14

**Introduction, and Theoretical Framework**

On November 8, 2016, the attention of not just the Western world, but the international community as well, shifted towards the United States of America. During those crucial hours, a remarkable political event unfolded, ultimately leading to one of the most astonishing and unforeseen electoral upsets in American history. A scenario that had hitherto been inconceivable began to materialise, as the Republican Party's candidate, Donald Trump, a New York entrepreneur, emerged victorious over his highly favoured opponent, Hillary Clinton, a symbol of the American establishment and a prominent figure in the Democratic Party. The US electorate had made its choice, ushering in a disruptive and transformative outcome. In a momentous and contentious collective decision, Americans bestowed the distinguished title of “leader of the free world” upon a figure who had been characterised by his opponents as unsuitable for the roles of president and commander-in-chief (Krieg & Diaz, 2016), and denounced as a racist (Lopez, 2016), misogynist, and sexist (Nelson, 2016). This result represented a profound and nearly inexplicable departure from the path undertaken just eight years prior, on November 4, 2008, when the Democratic candidate Barack Obama, a youthful Senator from Illinois, achieved an overwhelming triumph over his Republican counterpart, John McCain, becoming the first African American elected to the esteemed office of the United States Presidency. Obama himself, in a pre-electoral address to the nation, proclaimed, “no matter what happens, the sun will rise in the morning” (*Obama: No Matter What Happens, the Sun Will Rise in the Morning*, 2016), attempting to assuage concerns of those who feared that a conservative reaction would irrevocably jeopardize the destiny and aspirations of those who did not align with Donald Trump's ideologies, rhetoric, and desires. However, if the social scientist acknowledgesthat history transpires through a series of interconnected and contingent events (Sewell, 2005) and that events themselves bring about structural alterations (Sewell, 1996), in retrospect, it becomes apparent that Obama's claim was erroneous.

Indeed, while Trump's tenure has now concluded, the United States have experienced irrevocable changes since the 2016 election, exemplified by incidents such as the violent insurrection by fervent GOP supporters at the Capitol (Dalsheim & Starrett, 2021), as well as the unproven allegations of electoral fraud (Berlinski et al., 2023). Notably, it is the Republican Party that has undergone a substantial metamorphosis over the course of less than a decade. Previously endorsing the candidacy of a figure universally esteemed and admired like John McCain, who was eulogized as “the epitome of bravery” (McFadden, 2018) upon his passing, the party has since embraced a distinct persona that is largely disconnected from a considerable segment of the nation’s populace. This new profile has garnered explicit support from various white supremacist groups and far-right movements in America (Espinoza, 2021; Taylor, 2018; Von Mering & McCarty, 2013). Hence, it can be observed that between the electoral cycles of 2008 and 2016, a discernible transition towards radicalization is apparent within the Republican Party, and it is particularly evident in its official declarations and the defining attributes of its most prominent representatives. In my bachelor’s thesis (Guarnerio, 2021) I find that an analogous rightward shift also occurred within the realm of electoral programs. However, the ramifications of this transformation exhibit a strikingly greater level of intricacy than the perspectives and interpretations commonly promoted by numerous pundits, opinion leaders, and Western policymakers.

By making systematic qualitative comparisons between policy proposals contained in the 2008 (Republican Platform Committee, 2008) and 2016 (Republican Platform Committee, 2016) Republican platforms, pertaining to several of the most salient economic, and socio-cultural matters in the US context,I observethat the 2016 Republican Platform Committeeadopts a more aggressive attitude towards the Democratic Party, repeatedly portrayed as the representative *par excellence* of the established political order and the main culprit of the United States of America’s decline in power and prestige in the international arena. Furthermore, Republicans embraced more conservative positions on ethnic, religious, military, geopolitical and migration issues. On the other hand, the more the GOP’s political elaboration emphasises strictly material or economic topics, such as financial, monetary, labour, or infrastructural policies, the less of a deviation there is from the doctrines of liberalism and *laissez-faire*, which pervade the 2008 program and are traditionally associated with the Republican Party. Thus, my contribution reinforces the directions indicated by academic literature that focuses on the socio-cultural dimension of America’s right-wing extremism (Wilkinson, 2019; Williamson et al., 2011) while challenging those scholars who promotethe idea of an economically driven radicalisation, based on either an American version of welfare chauvinism (Komlos, 2018), or an ethnonationalist and protectionist reinterpretation of the GOP’s neoliberal workhorses (Post, 2017).

Despite its robust foundations, such analysis may be deemed as inconclusive, as it suffers from the intrinsic shortcomings of qualitative textual comparison, failing to captureelements that could instead be brought to light by a quantitative text analysis (Bernard & Ryan, 2015; Roberts, 2000). Moreover, my bachelor’s thesis does not encompass the key component of legislation, which is often presumed to be the primary manifestation of any electoral platform (Austen-Smith & Banks, 1988). To address these limitations, in this study I wish to delve into the following research question:

**RQ:** Did the Republican Party increase its legislativeactivities within the US Congress in tandem with its base and leaders’ radicalisation between 2008 and 2016? Were such initiatives more economically or socio-culturally oriented?

Empirical evidence in the substantive field of political science suggests that in the United States legislative agendas are sharply influenced by the platform of the President’s party in the short term, although this relationship is strongly shaped by issue-dependent differences, and fades over time (Fagan, 2018). Moreover, both Republicans and Democrats are affected by policy areas that the base and the overall public views as more important (Benefiel & Williams, 2019). In line with the literature and my bachelor thesis’ framework (Guarnerio, 2021), I expect the GOP’s parliamentary undertakings to be driven by the party’s polarisation, which in the timeframe of interest – i.e., 2008–2016 – was largely delimited to the socio-cultural dimension. More formally, I aim to probe two hypotheses:

**H1:** The Republican Party did not experience a notable increase in its legislative activities focused on economic issues within the US Congress between Obama's and Trump's first mandates in the Presidential office.

**H2:** The Republican Party witnessed a surge in its legislative activities pertaining to socio-cultural matters within the US Congress between Obama's and Trump's first mandates in the Presidential office.

To evaluate **H1** and **H2**, I deploy the computational tools of automated content analysis (ACA) because partisan legislative engagement in the US Congress is most accurately measured by the laws – i.e., the bills – each party introduces in the House of Representatives, and the Senate. This means that the social scientist must collect and examine thousands of proposed measures per Congressional term, a task that can most optimally be carried out with the ACA approach, by markedly reducing the costs of scrutinising large collections of texts (Grimmer & Stewart, 2013). This report is structured into three sections, each encompassing a different aspect related to my inquiry, following its overall workflow. In the next chapter, I thoroughly explain my analytic strategy, respectively focusing on the data collection procedures, and the training, validation, and testing of severalSupervised Machine Learning (SML) models geared towards solving the classification tasks of interest. Subsequently, I illustrate the statistical techniques I implement for the formal assessment of my hypotheses and present my study’s results. I conclude the paper by discussing the empirical evidence in comparison to my bachelor’s thesis’ findings, providing an answer to the **RQ**, and suggesting multiple directions for future research, to address my work’s limitations.

**Analytical Strategy: Data Collection**

I define legislative activities as the initial draft introductions of bills in the chambers of the US Congress– i.e., the House of Representatives, and the Senate. Importantly, I consider submissions of bills approved by the House of Representatives into the Senate as distinct parliamentary actions, even though the policy content may be identical, because my substantive interest concerns all partisan endeavours to enact electoral platforms. In line with the research’s theoretical framework, given that in the United States law-making agendas are shaped by President’s party’s platform prevalently in the short run (Fagan, 2018), I focus on the congressional terms correspondent to Obama's first and Trump's first mandates in the Presidential office – i.e., the 111th (2008 – 2010), and 115th (2016 – 2018) sessions. I gather such data from the US Congress’ application programming interface (API), the most suitable choice to inspect, retrieve, and store machine-readable information from the American government’s public collections. Appropriatelypre-processing the bills’ textual contents would be remarkably challenging due to their length and complexity, therefore I decide to fetch the plainer, yet highly informative bill summaries officially provided by the US institutions as features for the classification tasks of interest. Since the API endpoint dedicated to summaries is unavailable, it is necessary to iterate over each individual bill to gather its distinct identifiers and overview. Consequently, the data collection procedure becomes significantly time-consuming to execute, posing notable challenges in terms of efficiency and duration. Toavoid connectivity issues, I run all scripts on the Microsoft Azure ML cloud computing service. I obtain summaries for a total of 21820 bills – i.e., 13956 for the House of Representatives, and 7864 for the Senate. To enrich the accessible information, I acquire metadata concerning the policy area of each document and its sponsor's name, surname, state of election, and partisan affiliation, employing a comparable methodology. Ultimately, the only missing values are the policy areas for 76 bills, a shortcoming that does not substantially impact my analytic strategy. Key attributes regarding the fetched data and metadata are reported in Table 1.



Next, I randomly select 2200 documents for human coding – i.e., about 10% of the analytic sample at my disposal. In accordance with my bachelor’s thesis’ theoretical framework (Guarnerio, 2021), I conceptualise bills in the “Economic” category as legislative items whose summaries comprise policies, initiatives, or provisions that primarily focus on matters related to the economy, finance, trade, taxation, labour, and infrastructure investments. On the other hand, I define bills in the “Socio-Cultural” category as legislative items whose summaries encompasspolicies, initiatives, or provisions that mainly address social and cultural issues, and topics concern the overall well-being of the American society, such as social justice, cultural integration, diversity and inclusion, welfare programs and benefits, public and private healthcare, education, environmental protection, crime, gun control, immigration, and international relations. The classification steps are distinct, because a single record may be categorised as both “Economic” and “Socio-Cultural” in nature. Results of manual labelling are reported in Table 2. 1233 bills are classified as “Economic”, while 1371 are tagged as “Socio-Cultural”. It is understandablethat most labels for each category are positive, as their *a priori* definitions are broad and generic in essence.I appropriately account for class imbalance when training my Supervised Machine Learning classifiers by maximising the macro F1 performance metric, which averages each model’s accuracy on the individual categories independently of their frequency in training, validation, and testing datasets (De Angeli et al., 2022).



**Analytical Strategy: Supervised Machine Learning**

Since I must code the political contents of 19620 unlabelled documents of interest as economic, and socio-cultural, I face two highly domain-specific categorisation problems that are grounded on abstract, and latent concepts. My main goal is not to explore textual data to discover previously unknown patterns, but to test hypotheses concerning policy contents on which I hold robust theoretical expectations. Given these conditions, and in accordance with Boumans and Trilling’s (2016) recommendations, I opt for employing Supervised Machine Learning (SML) techniques, highly valuable for deductive labelling of extensive datasets. In other words, I aim to automatically replicate my human classifications, and construct a large-scale analytic sample, by training, testing, and applying a wide range of SML algorithms, including well-established approaches, such as the Bag-Of-Words (BOW) feature representation (Van Atteveldt et al., 2021), and Support Vector Machine modelling (Noble, 2006), alongside state-of-the-art transformers (Vaswani et al., 2017). To reduce noise within the retrieved summaries, which were likely web scraped by the US Congress’ API developers, I apply custom pre-processing and remove HTML tags and character escapes, following HaCohen-Kerner et al.’s(2020) methodological suggestions.



As a baseline, I carry out hyperparameter fine-tuning on several shallow machine learning classifiers of standard application (Xu et al., 2021) – i.e., Naïve-Bayes, Logistic, Support Vector Machine, and Random Forest algorithms (Van Atteveldt et al., 2021). Furthermore, I test two automatic NLTK vectorizers (Bird et al., 2009), respectively grounded on raw word counts, and term frequency–inverse document frequency (TF-IDF) weighting (Paik, 2013). Metrics measuring performance on the testing set**1** for the models that yield the best outcome in the post-training validation steps within this co-occurrence analysis framework are respectively reported in Tables 3 and 4. The most optimal classifier for identifying bill summaries as “Economic” or “Non-Economic” **2** is a SVM with the linear kernel trick (Wu et al., 2005), and an increased regularisation parameter – i.e., C = 10 – in combination with a TF-IDF vectorizer pruning words recurring in at least 75% of the documents contained in the training set. It is the only solution that achieves acceptable accuracy – i.e., 0.77 – while exhibiting robust performances across the board – i.e., no precision, or recall, is lower than 0.76. This aspect is crucial, because all the other candidates tend to artificially inflate the number of positive labels – i.e., show insufficient recalls for the “Non-Economic” category. Turning to the second classification task, the bestclassifier for tagging bill summaries as “Socio-Cultural” or “Non-Socio-Cultural” is a conventional Naïve-Bayes algorithm, in tandem with a raw word count vectorizer pruning words recurring in at least 50% of the documents contained in the training set, and removing NLTK’s default list of English stop-words (Van Atteveldt et al., 2021). This solution attainsan impressive overall accuracy – i.e., 0.82 – but its decisive shortcoming is the limited recall for the “Non-Socio-Cultural” group – i.e., 0.71 – yielding a harmonised F1 score of just 0.74. These figures indicate a tendency of unjustifiably augmentingthe number of positive labels, likely caused by the genericity and abstractness of the categories I deployedduring manual annotation, ultimately pointing towards the necessity of transcending BOW and co-occurrence analysis with deep learning approaches that take temporality and context into account.





Transformers represent the current state-of-the-art for textual feature representation and modelling. Unlike the traditional convolutional or recurrent neural networks, they are grounded on the so-called self-attention mechanism (Vaswani et al., 2017), which enables them to optimally handle intricate dependencies, precisely capture highly contextual information, and achieve improved efficiency in training and inference times. In this study, I implement two popular Bidirectional Encoder Representations from Transformers (BERT) architectures (Devlin et al., 2018) – i.e., base BERT, and the more domain-specific LEGAL-BERT (Chalkidis et al., 2020). The former is pre-trained for masked language modelling (MLM) and next sentence prediction (NSP) on a large corpus of English data in a self-supervised manner, whereas the latter utilises a web-scraped collection of English legal text encompassing several substantive fields – e.g., EU and UK legislation, EU and US court cases, and US contracts. Both models are fine-tuned to solve the classification tasks of interest within the Google CoLab cloud computing environment. Metrics measuring performance on the testing set for the BERT versions that yield the most promising predictions in the post-training steps are respectively reported in Tables 5 and 6**3**. The superior transformer for identifying bill summaries as “Economic” or “Non-Economic” is LEGAL-BERT, which attains an impressive overall accuracy – i.e., 0.82, markedly higher than the most optimal shallow learning solution– while maintaining the crucial property of exhibiting robust performances across the board – i.e., no precision, or recall, is lower than 0.80. This implies that LEGAL-BERT does not show any tendency of distorting the number of positive labels – i.e., displaying unsatisfactory recalls for the “Non-Economic” category. Turning to the second classification task, the superior transformer for tagging bill summaries as “Socio-Cultural” or “Non-Socio-Cultural” is base BERT, which achieves an outstanding accuracy – i.e., 0.89. Although it still manifests a tendency to disproportionately increase the number of positive predictions – i.e., the recall for the negative label is just 0.73 – the harmonised F1 score for the “Non-Socio-Cultural” class – i.e., 0.82 – indicates that this is a comparatively mitigated issuethan in the best shallow learning solution. Ultimately, it appears that discerning “Economic” between “Non-Economic” documents is more dependent on a deeper understanding of legal terminology, whereas distinguishing “Socio-Cultural” from “Non-Socio-Cultural” summaries is largely contingent on the comprehension of a more general language domain.



To conclude, I employ the fine-tuned LEGAL-BERT for predicting “Economic” versus “Non-Economic” labels, and the optimised base BERT for attributing “Socio-Cultural” versus “Non-Socio-Cultural” designations. The downstream tasks are loaded and executed on the Google CoLab cloud computing environment to ensure efficient computation. The classification results of the unlabelled data are presented in Table 7. Out of the remaining 19620 summaries, 11013 are tagged as “Economic” and 13598 as “Socio-Cultural”. The predominance of positive categories is expected and aligns with the prevalence observed in the labelled dataset. A visual inspection of the estimations further validates the performance metrics shown in Tables 5 and 6.



**1.** In accordance with the standard procedure of partitioning textual datasets into training, validation, and testing splits in a 60-20-20% ratio (Van Atteveldt et al., 2021), I allocate 1320 instances for training, 440 for validating, and 440 for testing purposes.

**2.** Due to random chance, a slight class imbalance issue arises in the test set, where 222 documents are classified as "Economic" and 218 are categorized as "Non-Economic." It should be noted that this condition is unrealistic since the positive labels are predominant in the original dataset. Therefore, the results presented in Tables 3 and 5 should be interpreted with caution.

**3.** For training the BERT models, a total of five training epochs are executed with a batch size of 8 documents. The models are trained using a standard learning rate of 5e-5, an L2 regularization weight decay of 0.01, and an optimiser with 200 warmup steps. It is important to note that these hyperparameters are not fine-tuned due to frequent issues encountered in GPU memory and RAM allocation within the Google CoLab environment.

**Exploratory Analysis, Statistical Analysis, and Results**

Before conducting a formal evaluation of **H1** and **H2**, I visualise the data enriched with Supervised Machine Learning Techniques. The focus is to illustrate the temporal distribution of bills classified as either Economic or Non-Economic, and Socio-Cultural or Non-Socio-Cultural, introduced in both chambers of the US Congress. Specifically, I confront the legislative activities of the two major parties, namely the Republicans and Democrats, during the first two years of President Obama's presidency, which corresponds to the 111th Congress (2008–2010), with the initial two years of President Trump's mandate, represented by the 115th Congress (2016–2018). This comparative exploration aims to provide a more nuanced understanding of the legislative agendas pursued by liberals and conservatives during their respective early terms.Categorical plots for Republicans are shown in Figures 1 and 2, while distributions for Democrats are displayed in Figures 3 and 4.

Immagine che contiene testo, schermata, diagramma, Diagramma

Descrizione generata automaticamente

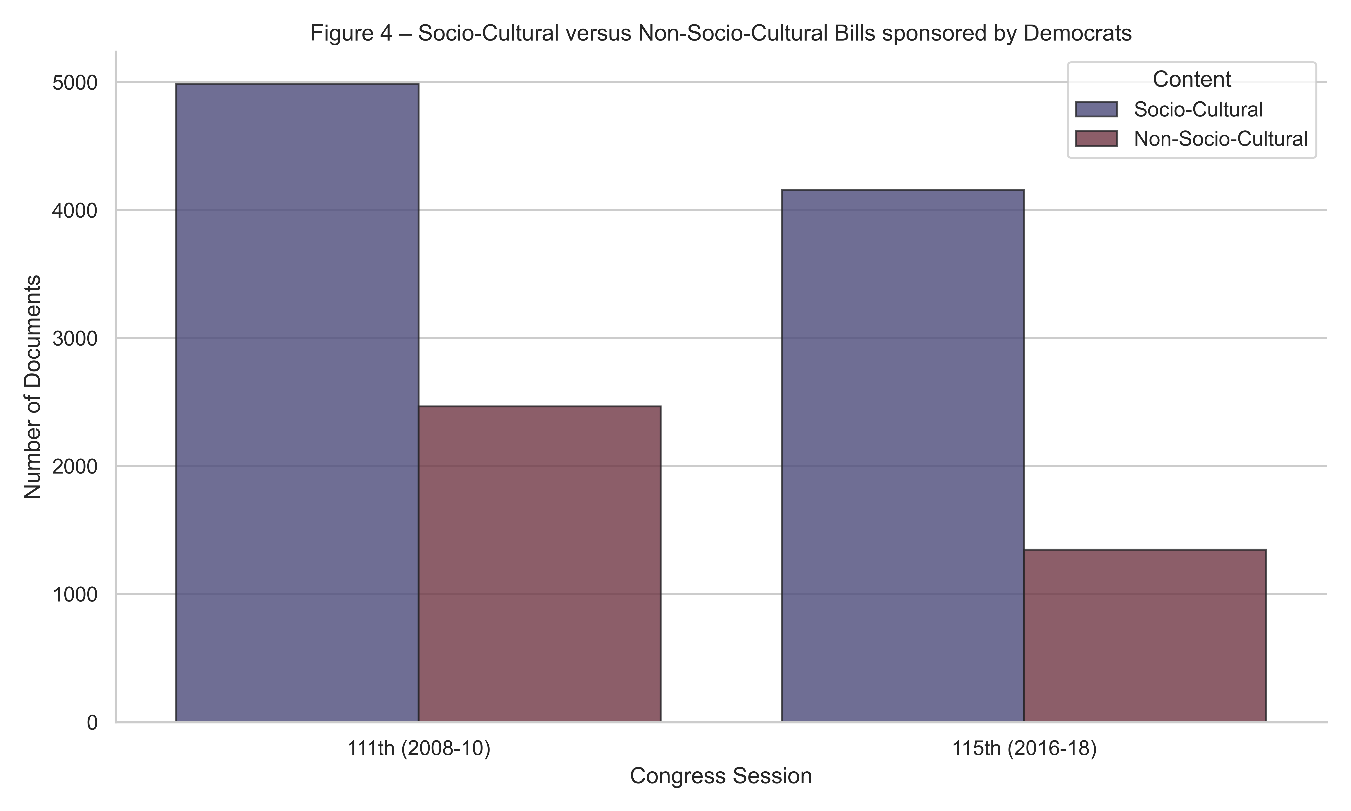
Immagine che contiene testo, schermata, diagramma, Diagramma

Descrizione generata automaticamente

In the 115th Congress, the Republicans sponsored a markedly higher volume of bill introductions, which aligns with my theoretical expectations. In the 111th Congress, the Republicans did not secure the presidency and held a minority representation in both the House of Representatives and the Senate. However, in the 115th Congress, they nominated their candidate as President and achieved majorities in both parliamentary chambers. The observed increase in the number of bills related to socio-cultural issues depicted in Figure 2 is noteworthy. Additionally, Figure 1 showsthat conservative legislative activity during Trump's presidency was not primarily driven by economic concerns. These findings lend support to my hypotheses. However, since the categories of interest are not mutually exclusive, it is necessary for me to carry out a statistical analysis to estimate and disentangle the main effects, specifically the overtime probabilities associated with introducing legislative documents focused on economic or socio-cultural issues.

Immagine che contiene testo, schermata, diagramma, Rettangolo

Descrizione generata automaticamente



Turning to the Democrats, in the 115th Congress, liberal sponsors introduced a slightly lower volume of bills compared to the 111th Congress, aligning with my theoretical expectations. During the 111th Congress, the Democrats not only secured the presidency but also achieved a majority in both the House of Representatives and the Senate. Conversely, in the 115th Congress, they did not elect the President and held minority representation in the two legislative chambers. Figure 4 illustrates how this shift had a consistent impact on socio-cultural concerns within parliamentary activities. Dynamics in the material dimension were more pronounced, as Figure 3 displays how there was an overtime decline in the prominence of economic legislation. It is plausible that this trend is attributable to the absence of budgetary and financial policies relating to public healthcare during the Trump administration, in contrast to the initial term of the Obama presidency. A comprehensive investigation of this proposition is left as a subject for future research.

To formally assess **H1** and **H2**, I conduct two fixed-effects OLS regressions to examine the likelihood of a bill sponsored by a Republican being categorized as Economic or Socio-Cultural, focusing on the effect of time. The independent variable of interest is a dummy indicating whether the bill was introduced in the 115th Congress – coded as 1 – or in the 111th Congress – coded as 0. I control for state-specific characteristics to account for the variations between legislators elected by different constituencies. In this analysis, I choose the US State of Florida as the reference category due to its historical significance as a swing state that is experiencing increasing conservatism among the electorate, driven by polarization on economic and socio-cultural issues (MacManus et al., 2005). The decision of employing linear probability models (LPMs) instead of logistic regression in this study is justified by the potential impact of omitted covariates on the estimates obtained through the latter approach, even when these covariates are unrelated to the predictors included in the specification. The existing literature emphasizes the challenges faced by social scientists in interpreting log-odds ratios or odds ratios as straightforward measures of effects and comparing them across different models. This difficulty arises due to the reflection of unobserved heterogeneity within the regression coefficients, due to mathematical properties inherent to the logistic function (Mood, 2010). Given the absence of a priori justification to suppose that the homoscedasticity assumption holds, I conduct White's tests for heteroscedasticity of residuals, instead of the conventional Breusch-Pagan approach (Breusch & Pagan, 1979). White's technique enjoys the advantage of detecting a broader range of heteroscedasticity forms (White, 1980). Both OLS models are subjected to these tests, yielding p-values that approach zero, providing compelling evidence to reject the null hypothesis of homoscedasticity of residuals. Consequently, I compute heteroscedasticity-robust standard errors with the “HC3” covariance matrix, following the methodological recommendations of Scott Long and Ervin (2000).



Results of the fixed-effects LPMs are respectively displayed in Tables 8 and 9. During the 115th Congress, bills introduced by Republican sponsors exhibited a marked decrease of 10.8% in the likelihood of being categorised as economic in nature. This negative effect remains statistically significant at the 95% confidence level even after accounting for state-specific variations among legislators representing different constituencies. Thus, my empirical evidence corroborates **H1**, as the GOP did not witness a substantial increase in its legislative activities focused on economic issues within the US Congress between President Obama's and President Trump's first terms in office. Conversely, bills introduced by Republican sponsors during the 115th Congress displayed a notable increase of 8.3% in the likelihood of being categorized as socio-cultural in nature. This positive effect remains statistically significant at the 95% confidence level, even when controlling for state-specific diversities among legislators elected from various constituencies. These findings provide empirical support for **H2**, as the Republican Party experienced a surge in its legislative activities pertaining to socio-cultural matters within the US Congress between President Obama's and President Trump's first terms in office. However, the adjusted R2 values for both LPMs are remarkably low, suggesting inadequate overall model fit. This indicates that the predictors included in the models explain only a small portion of the total variation in the outcome variables. Consequently, further research is warranted to explore additional potential confounders that may have a significant impact on the phenomenon of substantive interest.



**Discussion, Conclusions, and Limitations of the Study**

In this paper, I focus on the assessment of the over-time radicalisation of the Republican Party in the economic and socio-cultural dimensions. I expand on my bachelor’s thesis’ framework, which provides support for the academic literature that emphasises the importance of socio-cultural claims and issues in studying America’s right-wing extremist movements (Wilkinson, 2019; Williamson et al., 2011), while challenging interpretations grounded on economic factors and dynamics (Komlos, 2018; Post, 2017). By employing state-of-the-art Supervised Machine Learning approaches for automated content analysis of large-scale textual datasets, I evaluate the extent to which the partisan polarisation I detected in a qualitative study of the GOP’s 2008 and 2016 electoral platforms (Guarnerio, 2021) is reflected in conservatives’ short-term legislative activities within the US Congress. Findings derived from fixed-effects linear probability models provide empirical support for the hypotheses **H1** and **H2**. Between Obama’s and Trump's first terms in the Presidential office, the Republican Party did not experience a notable increase in its legislative activities devoted to economic issues yet witnessed a surge in its parliamentary initiatives pertaining to socio-cultural matters. As such, I identify evidence that the GOP intensified its legislativeendeavours within the US Congress in tandem with its base and leaders’ radicalisation between 2008 and 2016, with these undertakings being less economically and more socio-culturally oriented.

This research is built upon robust theoretical and methodological foundations, yet it is important to acknowledge several inherent shortcomings, which may affect the extent and generalizability of my conclusions. First, the data collection process could be expanded by taking more US Congress terms into consideration, or by directly web scraping each bill’s text, to provide a richer information pool as an input for Supervised Machine Learning techniques. However, scaling up document retrieval would be remarkably challenging since the US Congress’ application programming interface only allows for 1000 requests per hour, with thousands of parliamentary initiatives being introduced during every two-year session. Cloud computing services such as Microsoft Azure and Google Cloud would prove invaluable for this task, enabling the social scientist to achieve maximum efficiency and avoid connectivity issues by running all scripts on dedicated servers. Second, this study’s conceptualisation and operationalisation of the economic and socio-cultural dimensions in legislation is broad and generic in nature, leading to SML classifiers over-representing positive labels. Replicating training and analysis with differently devised categories could aid in testing the external validity of my findings.

Furthermore, it is crucial to note that manual coding was carried out by a single individual, increasing the risk of distorting the classification tasks with subjectivity and bias. The standard procedure involves a team of human annotators or experts, and empirical validation with inter-coder reliability measures (Lombard et al., 2002) such as Cohen’s kappa (Cohen, 1960) and Krippendorff’s alpha (Hayes & Krippendorff, 2007). Third, RAM, memory, and usage limits of Google CoLab’s free-to-use GPU accelerators prevent me from training and hyperparameter fine-tuning larger BERT architectures – e.g., the BERT large model (Devlin et al., 2018) – pointing towards the need of supplementary computational power. Fourth, extending the information collection process would open opportunities to estimate mixed-effects models, which account for the hierarchical structure of the data, with bills nested within Congress terms. However, the number of Level-1 units, representing individual documents, would considerably outnumber the Level-2 clusters, corresponding to Congress terms. Consequently, the social scientist would require utilising restricted maximum likelihood (REML) estimators for variance parameters, and the application of a t-distribution with a Satterthwaite (1946) or Kenward-Roger (1997) denominator-degrees-of-freedom (DDF) adjustment technique for statistical inference, as recommended by Elff et al. (2021). To conclude, future research could substantively focus on investigating how Democrats reacted to the GOP’s socio-cultural radicalisation, by constructing and implementing new political repertoires. Transcending the textual dimension of platforms and bills to analyse liberal and conservative US media content, building upon the existing literature’s suggestions (Martin & Yurukoglu, 2017), could aid in realising the untapped potential of computational methods within the field of communication science (Van Atteveldt & Peng, 2018).

**References**

Austen-Smith, D., & Banks, J. (1988). Elections, Coalitions, and Legislative Outcomes. *American Political Science Review*, *82*(2), 405–422. https://doi.org/10.2307/1957393

Benefiel, C., & Williams, C. J. (2019). Taking official positions: How public policy preferences influence the platforms of parties in the United States. *Electoral Studies*, *57*, 71–78. https://doi.org/10.1016/j.electstud.2018.11.001

Berlinski, N., Doyle, M., Guess, A. M., Levy, G., Lyons, B., Montgomery, J. M., Nyhan, B., & Reifler, J. (2023). The Effects of Unsubstantiated Claims of Voter Fraud on Confidence in Elections. *Journal of Experimental Political Science*, *10*(1), 34–49. https://doi.org/10.1017/XPS.2021.18

Bernard, H. R., & Ryan, G. (2015). Text Analysis: Qualitative and Quantitative Methods. In H. R. Bernard (Ed.), *Handbook of methods in cultural anthropology* (pp. 595–645). Rowman & Littlefield.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O’Reilly.

Boumans, J. W., & Trilling, D. (2016). Taking Stock of the Toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, *4*(1), 8–23. https://doi.org/10.1080/21670811.2015.1096598

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287. https://doi.org/10.2307/1911963

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The Muppets straight out of Law School. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2898–2904. https://doi.org/10.18653/v1/2020.findings-emnlp.261

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Dalsheim, J., & Starrett, G. (2021). Everything Possible and Nothing True: Notes on the Capitol Insurrection. *Anthropology Today*, *37*(2), 26–30. https://doi.org/10.1111/1467-8322.12645

De Angeli, K., Gao, S., Danciu, I., Durbin, E. B., Wu, X.-C., Stroup, A., Doherty, J., Schwartz, S., Wiggins, C., Damesyn, M., Coyle, L., Penberthy, L., Tourassi, G. D., & Yoon, H.-J. (2022). Class imbalance in out-of-distribution datasets: Improving the robustness of the TextCNN for the classification of rare cancer types. *Journal of Biomedical Informatics*, *125*, 103957. https://doi.org/10.1016/j.jbi.2021.103957

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. https://doi.org/10.48550/ARXIV.1810.04805

Espinoza, M. (2021). Donald Trump’s impact on the Republican Party. *Policy Studies*, *42*(5–6), 563–579. https://doi.org/10.1080/01442872.2021.1950667

Fagan, E. J. (2018). Marching Orders? U.S. Party Platforms and Legislative Agenda Setting 1948–2014. *Political Research Quarterly*, *71*(4), 949–959. https://doi.org/10.1177/1065912918772681

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Guarnerio, M. (2021). *Rumore Bianco—La radicalizzazione del Partito Repubblicano nel confronto tra due programmi elettorali (2008—2016)*. Università degli Studi di Trento.

HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLOS ONE*, *15*(5), e0232525. https://doi.org/10.1371/journal.pone.0232525

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, *1*(1), 77–89. https://doi.org/10.1080/19312450709336664

Kenward, M. G., & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, *53*(3), 983. https://doi.org/10.2307/2533558

Komlos, J. (2018). The Economic Roots of the Rise of Trumpism. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3144271

Krieg, G., & Diaz, D. (2016, October 10). Donald Trump vs. Hillary Clinton II: The nastiest lines. *CNN*. https://cnn.it/45DM5WA

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research*, *28*(4), 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x

Long, J. S., & Ervin, L. H. (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician*, *54*(3), 217–224. https://doi.org/10.1080/00031305.2000.10474549

Lopez, G. (2016, July 25). Donald Trump’s long history of racism, from the 1970s to 2020. *Vox*. http://bit.ly/Trump\_Racism\_Vox

MacManus, S. A., Everett, C., Quecan, A., & Penberthy, B. L. (2005). Florida: The South’s Premier Battleground State. *American Review of Politics*, *26*, 155–184. https://doi.org/10.15763/issn.2374-7781.2005.26.0.155-184

Martin, G. J., & Yurukoglu, A. (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review*, *107*(9), 2565–2599. https://doi.org/10.1257/aer.20160812

McFadden, R. D. (2018, August 25). John McCain, War Hero, Senator, Presidential Contender, Dies at 81. *The New York Times*. https://nyti.ms/3C5kgc5

Mood, C. (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review*, *26*(1), 67–82. https://doi.org/10.1093/esr/jcp006

Nelson, L. (2016, October 8). Donald Trump’s history of misogyny, sexism, and harassment: A comprehensive review. *Vox*. http://bit.ly/Trump\_Sexism\_Vox

Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565

*Obama: No matter what happens, the sun will rise in the morning*. (2016, November 9). [YouTube]. CNN. http://bit.ly/The\_Sun\_Will\_Rise

Paik, J. H. (2013). A novel TF-IDF weighting scheme for effective ranking. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 343–352. https://doi.org/10.1145/2484028.2484070

Post, C. (2017). The roots of Trumpism. *Cultural Dynamics*, *29*(1–2), 100–108. https://doi.org/10.1177/0921374017709229

Republican Platform Committee. (2008). *2008 Republican Platform*. 2008 Republican National Convention, Saint Paul, Minnesota. http://bit.ly/GOP\_Platform\_2008

Republican Platform Committee. (2016). *Republican Platform 2016*. 2016 Republican National Convention, Cleveland, Ohio. http://bit.ly/GOP\_Platform\_2016

Roberts, C. W. (2000). A Conceptual Framework for Quantitative Text Analysis. *Quality and Quantity*, *34*(3), 259–274. https://doi.org/10.1023/A:1004780007748

Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, *2*(6), 110. https://doi.org/10.2307/3002019

Sewell, W. H. (1996). Historical events as transformations of structures: Inventing revolution at the Bastille. *Theory and Society*, *25*(6), 841–881. https://doi.org/10.1007/BF00159818

Sewell, W. H. (2005). *Logics of history: Social theory and social transformation*. University of Chicago Press.

Taylor, K.-Y. (2018). The White Power Presidency: Race and Class in the Trump Era: NPS Plenary Lecture, APSA 2017, San Francisco, CA. *New Political Science*, *40*(1), 103–112. https://doi.org/10.1080/07393148.2018.1420555

Van Atteveldt, W., & Peng, T.-Q. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, *12*(2–3), 81–92. https://doi.org/10.1080/19312458.2018.1458084

Van Atteveldt, W., Trilling, D., & Arcíla, C. (2021). *Computational analysis of communication: A practical introduction to the analysis of texts, networks, and images with code examples in Python and R*. John Wiley & Sons.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. In U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, R. Fergus, S. V. N. Vishwanathan, R. Garnett, & Neural Information Processing Systems Foundation (Eds.), *Advances in neural information processing systems 30: 31st Annual Conference on Neural Information Processing Systems (NIPS 2017): Long Beach, California, USA, 4-9 December 2017*. Curran Associates, Inc.

Von Mering, S., & McCarty, T. W. (Eds.). (2013). *Right-wing radicalism today: Perspectives from Europe and the US*. Routledge.

Wilkinson, W. (2019). *The density divide: Urbanization, polarization, and populist backlash.* Niskanen Center. http://bit.ly/Density\_Divide

Williamson, V., Skocpol, T., & Coggin, J. (2011). The Tea Party and the Remaking of Republican Conservatism. *Perspectives on Politics*, *9*(1), 25–43. https://doi.org/10.1017/S153759271000407X

Wu, G., Chang, E. Y., & Panda, N. (2005). Formulating distance functions via the kernel trick. *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 703–709. https://doi.org/10.1145/1081870.1081960

Xu, Y., Zhou, Y., Sekula, P., & Ding, L. (2021). Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment*, *6*, 100045. https://doi.org/10.1016/j.dibe.2021.100045