THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE

Measuring changes in Vatican social policy from Papal documents

LSE Research Online URL for this paper: http://eprints.lse.ac.uk/101437/

Version: Published Version

Monograph:

Grzymala-Busse, Anna and Miller, Blake Andrew Phillip (2017) Measuring changes in Vatican social policy from Papal documents. . University of Michigan, Department of Statistics, Ann Arbor, Michigan.

lseresearchonline@lse.ac.uk
https://eprints.lse.ac.uk/

Measuring Changes in Vatican Social Policy from Papal Documents[*]

Anna Grzymala-Busse[†]   Blake A. Miller[‡]

April 4, 2017

This is an early draft. Findings presented are from a sample of cleaned and pre-processed documents from the larger corpus. Further drafts will make use of the entire corpus, not just a sample.

Abstract

There is much debate about how, when, and why the Roman Catholic Church has changed its emphasis on secular states and social issues. Unfortunately, large-scale, systematic research on papal policy has not yet been done, despite widely available papal documents on the subject. Using automated multilingual topic modeling and other methods of automated text analysis, we analyze a large multilingual corpus of papal encyclicals, bulls and other documents spanning the past 2000 years. We aim to use this corpus to better understand the evolving posture of the Church with respect to modern states and social policy. We adjudicate between two competing propositions, 1) that the Church's views change in response to doctrinal developments only (the naïve doctrine hypothesis), and 2) that the Church emphasizes social policy when it feels threatened, as a way of reasserting its moral authority. Additionally, we investigate the hypothesis that Church and state only became distinct concepts in modern history. This work, aside from adjudicating major debates surrounding the Church's emphases on social policy in secular states, also introduces a new method of topic modeling using multilingual corpora.

1

Introduction

How do religious groups change their doctrine? Specifically, how have they interpreted the secular state, and how have these views changed? Modern religions do not shy from political engagement: conservative Evangelicals have used abortion as a litmus test for judicial appointees in the United States, extremist groups have justified their military campaigns in the name of Islam, the Roman Catholic Church continues to control hospitals, charities, and welfare institutions in several countries, and ultra-Orthodox Jews have obtained policy concessions and financial subsidies from the secular state in both Israel and in the United States. Yet for all their modern encounters with the state, we know less about how religious groups have conceptualized the state earlier in history, or how these views have changed over time.

Religious views on the state matter in several ways. First, doctrine is unfalsifiable: no worldly evidence can overcome its claims. Since it is a belief system that cannot be disconfirmed, the claims of religion on politics can be absolute and irrefutable. Second, unlike other ideologies, doctrine provides divine sanction (and otherworldly rewards) for human action, making it a particularly compelling set of justifications and motivations for human behavior. Third, religious doctrine serves as a powerful template and a source of common knowledge (Chwe, 1994, 2013): a focal point and a framing device, a common understanding of both what others know and how they understand certain policy problems or challenges. The political views of the religious can change, with the Progressivist views of American evangelicals on Prohibition, populism, and abolition shifting to their contemporary conservatism and emphases on abortion, reproductive politics, and family. Yet, even as interpretations changed, the power of religious justification did not, sanctifying the goals and edifying the adherents. Doctrine can thus influence both the choice of institutions and the long-term outcomes that result. Woodberry (2011) argues that Protestant churches and missionaries promoted mass literacy, printing, and education, in an emphasis on the ability to read and to access the word of God directly. These in turn spurred social mobility, opportunities for women, and long-term health improvements, among other favourable outcomes. Dominant religious doctrines shaped would-be state institutions even before the state was founded. For example, Lutheran, Calvinist, and Catholic doctrine each views the source of poverty differently and accordingly, each

Church imposed different regimes of poverty assistance in early modern Europe. (S. Kahl, 2005; B. Kahl, 2014). Similarly, Church-State conflict in the nineteenth century, won by the secular liberal forces of the nation state, led to robust provisioning of the welfare state (Morgan, 2009).

As a result, religious views on the state can be enormously important both for the adherents and for the state itself. Since the institutional templates of the modern state first arose in Western Christendom, we investigate the Western Christian Church and its main successor after the Reformation, the Roman Catholic Church, the dominant religious denomination in Europe. Catholic Popes crowned monarch after monarch, and its doctrine of natural law mandated that God-given laws trump any man-made ones. The Church hired mercenaries, took sides in political alliances, and blessed both wars and peace treaties. It is deeply woven into the fabric of European history, its politics, the rise and development of the state, and the social policies that this state would eventually produce.

How, then, has the Roman Catholic Church viewed the state? To address this question, we examine three different sets of propositions: one that predicts that there was little differentiation between state and church until the modern era, one that expects initial doctrinal statements as holding constant, and one that views the church response as highly contingent and dynamic, dependent on the threat faced by the church. We systematically analyze the texts of Church documents and pronouncements over the last 2000 years to adjudicate between these hypotheses. Because our data are not yet fully processed, we cannot yet make any claims in support of any of these competing propositions. Instead, we outline our methods and report some interesting patterns found in our early exploration of this newly created dataset.

Explaining Religious Views

Religions are necessarily oriented toward the supernatural, the world beyond our temporal and physical context. Their very concern with the divine means that they are not wholly of this world. Much of religious philosophy and theology focuses on the nature of the divine, its demands and its teachings. But religions are not simply isolated commu-

nities of the faithful: they are part of a broader society and state that surrounds, defines, and even challenges them. The Roman Catholic Church thus confronted (and participated in) war, monarchical struggles, the demands of groups such as peasants, the rise of new technologies (such as the printing press, mechanization of production, or the Internet) and new ideologies (capitalism, communism, and so on.) It also responded: directly with exhortations for the faithful to obey specific rulers and support certain policies, and indirectly with a broader philosophy of the worldly state and society. The question then is how these views of the state have changed, if at all, and how.

In one scholarly view, the question is unintelligible, because it is ahistorical. In this account, the state is simply not a coherent concept in either political thought or in historical relationships, until relatively late, the late 1700s. No concept of the secular, or the distinction between church and state, arises until the nation-state forces the distinction.[1] Before then, neither religious thinkers nor political sovereigns saw a conceptual separation between state and religion, or between secular and religious authority (Armstrong, 2014; Cavanaugh, 2009). Other accounts point out that the lengthy medieval debates about the differences between human and divine law, and the distinctions between earthly and divine authority mean that a concept of the state was already present. In fact, Strayer's classic account posits that the Investiture Conflict, and Gregory VII's sharp assertion of church autonomy led to the rise of a modern state administration, with the church insisting that "justice was the essential attribute of secular rulers" (Strayer, 1970). By 1300, then, not only was a modern state administration arising, but so was the philosophical distinction. We should thus see minimal references to the state or the sovereign until either the 18th century (if Armstrong is correct) or the 14th (if Strayer is right), but once these begin, they should monotonically increase, given the predominance of the secular state in modern politics.

---

[1]As a result, Armstrong both argues that state and religion are indistinguishable: and that the secular state is responsible for violence. Cavanaugh does offer criteria for distinguishing the role of religion; a conflict is religious if: (1) combatants oppose each other on the basis of religious difference; (2) the primary cause is religious rather than political, economic, or social; (3) religious causes must be analytically separable from these other potential causes; and (4) the rise of the modern state can be eliminated as a cause (Cavanaugh, 2009, 141-2). Yet the second standard for distinguishing a religious cause is that religion must be a cause (2). This is less than helpful empirically. Moreover, Cavanaugh argues that the religious participation and legitimization of Yugoslav civil wars does not provide 'serious warrant for attributing violence to a sui generis impulse called religion' (Cavanaugh, 2009, 48). Yet the combatants often identified themselves by their religious affiliation (and more importantly, targeted others on that basis), and religious actors repeatedly motivated and justified the violence (Perica, 2002).

In another view, which we label the "naïve doctrine" hypothesis, state and church exist as separate entities in Christian theology from the beginning. This hypothesis finds its roots in Jesus's enjoinder that the faithful should pay taxes to the state, and should "render to Caesar the things that are Caesar's, and to God the things that are God's" (King James Bible, n.d., Mark 12:17). This is part of a broader tradition that views religions as possessing inherent doctrinal and political characteristics that change little over time as they are grounded in initial doctrinal commitments. For example, several scholars portray Islam and other denominations as intrinsically and unchangingly peaceful (Armstrong, 2014; R. Esposito, 2010; J. L. Esposito & Mogahed, 2007), but see Fish (2011) for a trenchant critique. We should thus observe a relatively constant and consistent set of emphases.

In a third possibility, the Christian Church emphasizes the notion of secular states and social issues when it is under threat. Rather than the doctrine either staying constant or gradually evolving, the Church responds to its strategic context. Thus, religious competition, in the form of the Protestant Reformation, for example, should lead the Church to more clearly distinguish itself and its theological "offer," and emphasize doctrinal differences with the Protestant challengers, rather than focus on the state or secular politics. This is in keeping with the "political economy of religion"approach, which views religious denominations as very responsive to competitive pressures (Chaves & Cann, 1992; Finke & Stark, 1992; Iannaccone, 1998; Gill, 2001; Clark, 2010). Where they face a threat to their hold on their congregations, they will respond with new doctrinal and theological offerings in an attempt to reinvigorate and attract adherents. The prediction, then, is that rather than a monotonic progression towards differentiation from the state and emphases on social policy, we will see periods of minimal change punctuated by decreases and increases in focus that correspond to competitive or exogenous threats to the Church.

This paper thus attempts to adjudicate between three competing propositions regarding the evolving posture of the Church with respect to modern states and social policy: a) that the church's views change in response to doctrinal developments only (the naïve doctrine hypothesis), b) that the church emphasizes these when it feels threatened, as a way of reasserting its moral authority, or c) that the church only differentiates between itself and other actors very late in the game, so both a) and b) are moot. To accomplish

this aim, we analyze a large multilingual corpus of papal encyclicals, bulls and other documents spanning the past 2000 years. These are the formal pronouncements of the Church hierarchy, which reflect the particular concerns and emphases of the global Church.

Data and Methods

Background on Papal Documents

Popes choose the form, subject, and audience of their official addresses. Accordingly, papal documents reflect a given Pope's priorities in doctrine, the institutional church, or social thought. They reflect his (and the Vatican's) views on what is salient and critical to address, and have generated considerable interest and attention once made public. They thus both reflect broader social, religious, and political trends–and help to shape what the faithful are to make of them. While no definitive compendium of papal encyclicals or other documents exists (there is no official corpus as there is with canon law), there is wide consensus on the most significant works, and these have been collected in the Vatican archives.

Papal documents fall into several categories. The highest and most binding of these are Apostolic Exhortations. These decrees are public in nature, and address solemn and significant church matters, whether doctrine, teachings, or law. Examples here include Ineffabilis Deus, which declared the Immaculate Conception of Mary, by Pius IX in 1854 or Ut Sit, the 1982 Apostolic Exhortation that raised the controversial and secretive lay Catholic order of Opus Dei to the status of a personal prelature by John Paul II.

Below the exhortations are papal encyclicals, which are official letters addressed to either all bishops and clergy or a subset of them. Some of the most significant statements of Catholic social thought and opinion on social matters have been issued as encyclicals, including Rerum Novarum (1891, on the rights and responsibilities of labor, trade unions, and management), Quadragesimo Anno (1931, on the ethical implications of capitalism), Mit brennender Sorge (1937, smuggled into Nazi Germany to condemn the idolatry of the "Reich Government" and its Neopaganism) or (Humanae Vitae (1968, which prohibited what the Church considers artificial contraception). Any papal document can be denoted

Motu Proprio, which indicates it comes from the Pope's own initiative and not in response to a specific individual or request.

Apostolic letters and exhortations generally address matters of lesser solemnity than encyclicals, usually having to do with canon law and internal church rules. Especially significant letters had a leaden Papal seal attached to them, and were known as Bulls. Those simply sealed with wax were known as Briefs, and generally referred to matters of lower importance. Several famous Bulls attempted to protect Jews (Sicut Judaeis, first issued in 1120 and repeatedly reaffirmed), excommunicated Martin Luther, approved the founding of the Jesuits, and established the Gregorian calendar.

All of these documents carry with them the authority of the Pope, who himself is the highest authority within the Church. This is not to say that all these documents are declared to be infallible statements: the confusion arises because only some documents are delivered "Ex Cathedra," which endows them with full and infallible authority. Such designation is rare, and no complete list is available, but perhaps the clearest and most significant modern example was a 1950 Apostolic Constitution declaring the dogma of the Assumption of the Virgin Mary (Munificentissumus Deus.)

Data

Our data consist of a collection of over 20,000 apostolic letters, papal speeches, encyclicals, bulls, letters, homilies, motu proprio documents, and other papal or Vatican documents from the first century CE to the modern era. (Popes also hold audiences, deliver homilies at masses and other services, give informal discourses, issue common declarations with other religious leaders and write letters to specific individuals. These are not systematically included here.)

Most documents in the corpus are authored by popes or Vatican leadership and are in several languages. Almost all documents are originally written in or translated into Latin, English, French, Spanish, and Italian. We obtained these documents through two main online libraries, G Catholic, and Documenta Catholica Omnia,the official Vatican website, and several smaller publicly available document collections.

Cleaning and Extracting the Data

Because these documents come from several sources, preparing and cleaning the data took considerable effort. Many documents are simply image scans of older documents that have not been converted to text. To process these data, we used optical character recognition (OCR) technology to extract text from documents using open source tools in each language (Smith, 2007).[2]

Pre-processing the Data

Once converted to plain text, we processed each document using natural language processing tools[3] for the languages represented in the corpus (English, French, Italian, Latin, Spanish). First we split the document into a list of its constituent words, an essential processing step for most automated text analysis methods. We then used a process called stemming to remove inflections from each of these words, thus unifying word features such as 'run,' 'ran,' and 'running.' Finally, we removed stopwords, which are uninformative, low-information words such as 'the' and 'and' from each document.

Now, we count the occurrence of resulting stemmed tokens representing each document. This results in a document term matrix of word counts where each column represents a word or a pair of co-occurring words (bigrams) in the vocabulary of the corpus, and each row represents a document in the corpus. These common pre-processing steps are frequently used in text analysis in political science (Grimmer & Stewart, 2013).

Though at this point we have converted our documents into numeric data in the form of a large, sparse matrix, we hope to remove noise and select word features to avoid over-fitting. This is accomplished by simply removing rare and common words[4]. Finally,

---

[2]Our efforts to digitize much of our text was made possible by an open source Latin OCR engine that can be found here: https://ryanfb.github.io/latinocr/. Unfortunately, many of these documents use archaic fonts, or are not high-quality scans, so there are many errors in the OCR. With all languages but Latin, this can be partially mitigated with spelling correction software, which we plan on using as we further clean data. In the case of Latin, we must build a custom spelling correction model. This spelling correction model is a noisy channel model that is trained on all professionally-digitized Latin text in the database, the Latin Vulgate bible, and a selection of Latin Wikipedia pages.

[3]We used the Natural Language Toolkit in Python for stemming and stopword removal in all languages but Latin. More information on NLTK can be found here: http://www.nltk.org. For Latin, we used the Classical Language Toolkit (CLTK) for stemming and stopword removal. More information on CLTK can be found at http://cltk.org

[4]A threshold for rare and common words is learned using a randomized search of several thousand possible models (Bergstra & Bengio, 2012).

we adjusted raw word counts with tf-idf weights.[5] The result is a sparse matrix of word weights for each document.

To further reduce noise and unify similar features such as 'pope' and 'pontiff,' we normalize the data and perform singular value decomposition (SVD) on the data matrix, a process known as latent semantic analysis (LSA) (Furnas et al., 1988). This process uncovers latent features in the data, ordered by importance, in a process similar to principal component analysis (PCA). To remove noise, we simply truncate the data matrix, resulting in a lower-dimensional representation of the data that preserves features of maximal variance.[6] After LSA is performed, we are ready for classification or topic modeling.

Categorization

Figure 1: Top Level Topics and Subtopics

We adopt a hierarchical topic categorization scheme, with multiple subtopics within three broad topic categories: doctrine, social issues and politics, and the institutional church.

Sub-topics fall under three main topics: religious doctrine, the institutional church, and church views on morality, politics, and social issues. The first category, religious doctrine, includes topics such as theology and religious competitors (threats). It comprises both church efforts to define the core elements of faith and Catholic beliefs, and effort to distinguish these beliefs from other religions and/ or competitors. Documents outlining what constitutes heresy or apostasy, as well as religious deviancy, all fall into this category.

---

[5]Tf-idf weights increase the weight of words that are in fewer documents and decreases the weight of words that are shared across several documents (Jurafsky & Martin, 2014). This produces better classification results.

[6]The number of features to keep is tuned by randomized hyperparameter search.

In many ways, this is the least worldly and most self-referential of the three categories, focusing as it does on the core elements of belief and doctrine.

The second category is the institutional church. This comprises the church as a human institution on earth, and includes church chaplaincies and outreach efforts, the church hierarchy itself, church meetings such as Vatican II or the various councils, the various religious orders, and the changing relationship of the church to other faiths and to the laity (the adherents.)

The Church has always lived in the world, and given its initial views on natural law as superseding human law, has formulated a wide variety of pronouncements, opinions, and directives regarding what it considers as domains falling under its purview, such as morality, family, and social justice. Accordingly, the third category comprises these views. The topics here range from bioethics to just war to culture, education, and family. Church opinions regarding law, economics, and just governance also fall into this category: what they all share in common is that the Church views these both as doctrinal directives for the faithful, and as its moral responsibility towards all human beings.

Supervised Topic Models

Figure 2: One-Versus-Rest Classifiers
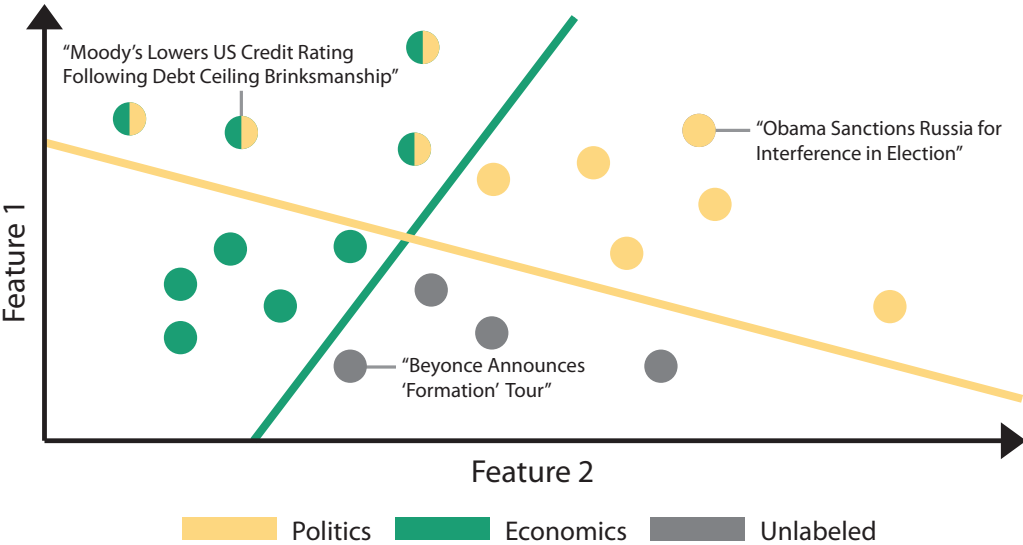
With two features, the lines colored lines represent hyperplanes for each topic class. A separate SVM for each class learns the boundary between the 'one' class, and the 'rest' of the labeled data points. Documents are represented as points, and the class of each document is represented by the color of the points.

Our topic models are supervised, meaning that we label documents according to pre-defined categories, rather than identifying latent topics from unlabeled documents. Supervised topic models are a better fit for our research needs because we want to impose a distinction between issues related to secular authority and spiritual authority. We also believe that supervision can aid, though not fully mitigate the problem of non-stationarity of languages represented in the corpus, which have likely evolved significantly over the two millenia represented in our data.

Because only a subset of our data is labeled, we must use statistical machine learning techniques to infer the topic of the thousands of unlabeled documents in our corpus. We impute topic labels for a subset of documents using document tags from around 3000 documents in the Catholic document archive, GCatholic.org. Though we currently use only these tag-imputed documents as a training set, we plan on hand labeling a stratified random sample of documents by pope and century to provide more temporal and linguistic representation in our training set.

We use the labeled documents to predict the topic membership of the approximately 20,000 remaining documents. Our algorithm of choice is a one-versus rest (OVR) multi-label support vector machine (SVM) classifier. This classifier learns a decision boundary by finding a vector that maximizes the margin between labeled observations from each class.

This classifier allows the outcome variable to be any combination of the unordered set of potential outcomes for both the broad and specific categories. A one-versus-rest SVM classifier is actually a combination of multiple binary classifiers, one for each class. Each classifier treats the members of it's class as a 1 and all other observations as -1. In this way, the set of topics of which a document is a member is determined by the decisions of component SVM classifiers for each of the 20 subtopics in our data.

For each of these topic classes, a class separating hyperplane is learned by optimizing the following quadratic equation. Because it has linear inequality constraints, this is a convex optimization problem (Friedman, Hastie, & Tibshirani, 2001, 417). An optimal hyperplane F is learned from training data $(x_1, y_1), \cdots, (x_n, y_n)$ with $y_i \in \{-1, 1\}$ giving

us weights β:

$$\min \|\beta\| \text{ subject to } \begin{cases} y_i(x_i'\beta) \geq 1 - \xi_i \ \forall i \\ \\ \xi_i \geq 0, \ \sum \xi_i \leq C \end{cases} \tag{1}$$

Here ξ_i allows for slack, and is the proportional amount that a data point falls on the wrong side of the class-separating hyperplane. Here, the cost parameter C represents a total budget for how much slack is allowed. C, for our purposes, is tuned via randomized hyperparameter search (Bergstra & Bengio, 2012).

Optimizing the function above gives us a hyperplane $F \subset \mathbb{R}^p$:

$$F = \{x : \beta_0 + x'\beta = 0\} \tag{2}$$

The classifier's decision is thus:

$$c(x) = sign\left(x'\beta + \beta_0\right) \tag{3}$$

More simply, our classifier classifies an out-of-sample document based on which side of the hyperplane the document falls. A graphical illustration of the topic classifier can be seen in Figure 2.

Since we have a separate classifier for each language, for each document, we create an ensemble classifier using all the languages a document is translated into. The topic of each document is decided by a 'vote' cast by the classifier for each language it is translated into:

$$c_i(x) = mode\left\{c_{i,en}(x), c_{i,es}(x), c_{i,fr}(x), c_{i,it}(x), c_{i,la}(x)\right\} \tag{4}$$

For example, if a document is translated into English, French, and Latin, our classification decision for the "Liturgy and Devotions" topic might look like:

$$c_i(x) = mode\left\{c_{i,en}(x), c_{i,fr}(x), c_{i,la}(x)\right\} = mode\left\{1, -1, 1\right\} = 1$$

Table 1: *Topic Classifier Performance*

Language	Micro-Avg F1	Macro-Avg F1	Weighted F1
English	0.6543	0.5716	0.6765
French	0.6907	0.5684	0.7111
Italian	0.6044	0.4952	0.6445
Latin	0.5210	0.2069	0.6245
Spanish	0.6400	0.5327	0.6663

Classifier Performance

Our classifier performance is acceptable, but because we are only using a subset of the labeled data available to us, there are several topic classes that have very few training examples, so our misclassification error is high for those class-specific component classifiers. Below we report the micro-averaged, macro-averaged and weighted F1 scores for the classifier. These are 3 different methods for averaging topic class' F1 scores (the harmonic mean of precision and recall). For more on how we calculate these performance metrics, see the appendix. These difference in micro, macro, and weighted F1 scores indicates that performance can likely be improved by adding labels to the under-represented classes in our training data.

Findings

Figure 3: Top Level Topic Proportions by Century
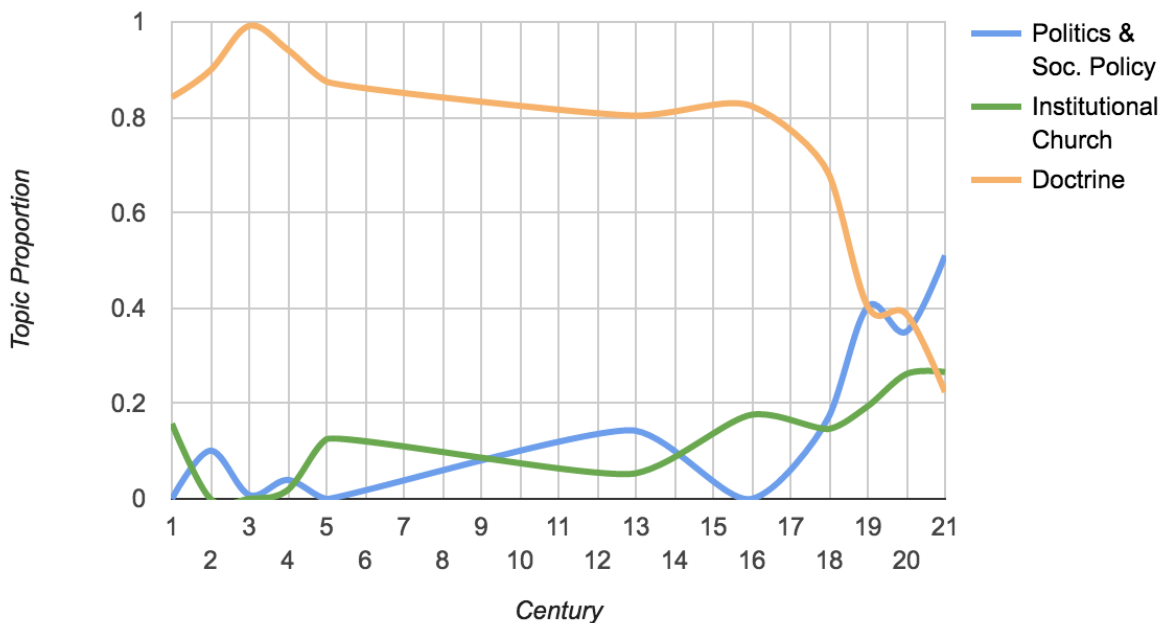
Topic proportions as predicted by our topic model can be seen in Figure 3 by century, in Figure 4 by pope, and Figure 5 by language.

We find evidence of an increase in mentions of politics and social policy over the 11-13th centuries, consistent with the findings of Blaydes, Grimmer, & McQueen (2013). Our findings are consistent with the notion that as the concept of the state became more philosophically developed, and states themselves became more autonomous and powerful entities, it became more salient to the Church. In other words, a better-defined conceptualization of the state made it more distinct from the Church. The differentiation and growth in central state structures, such as the rise of chanceries, tax collection agencies, and treasuries made these distinctions manifest. (Cavanaugh, 2009; Strayer, 1970). Note that while this timing is in keeping with Strayer's argument, it predates the predictions of Armstrong's view that church and state only become distinct concepts in the 18th century.

We also find evidence that the Roman Catholic Church began to pay more attention to politics and social policy after the 16th century. Conversely, the Church makes relatively fewer mentions of doctrine itself. This development does not necessarily occur because doctrinal questions were settled: the Protestant Reformation and the constant challenges to core Catholic doctrine that resulted was only taking off in the late 16th century. The Protestant Reformation, traditionally dated to Martin Luther's nailing of the 95 theses in 1517, was *the* challenge to Catholic orthodoxy: it threatened the Church's institutional prerogatives (the selling of indulgences, etc), the Church's emphasis on the mediation of religious experience by priests and clergy, and its view of salvation as earned by good works. Doctrinal controversies and debates erupted, whether in response to the Protestant challenge or to its Catholic response, the Counter-Reformation, led by the new order of Jesuits (doctrinaires par excellence.) Nonetheless, rather than responding to the Protestant threat with greater attention to doctrine and attempts to make the Catholic religious offering more distinct, as the "political economy of religion" might predict, the Church responded with greater relative attention to *politics*. This is consistent with the notion that the Church saw the religious competition as a challenge to its political status, and saw the solution as resting with princes and sovereigns, rather than with believers and their shifting loyalties. In contrast, in modern accounts of Church response to religious

competition, we see the Church paying greater attention tot the needs and demands of the local faithful, as in the case of late-20th century Mexico (**?**).

Significantly, the 16th century is also a period marked by a significant rise in the claims for state sovereignty and self-definition, which consist not only of the continued centralization and reassertion of the state as in the medieval period, but also of new and explicit political settlements and claims that affect the Church itself. This timing predates the predictions. For example, the Treaty of Augsburg in 1555 established the principle of "cuius regio, eius religio," (whose region, his religion), allowing rulers to choose whether their countries would officially adhere to Catholicism or Lutheranism. The Thirty Years' War raged across Europe from 1618-1648, with religious competition as its initial justification, and the millions of casualties and deaths as its result. While the religious nature of the conflict remains controversial (Armstrong, 2014; Cavanaugh, 2009), its consequences are far more established. The Peace of Westphalia (1648) further allowed Calvinism as a state religion, and established a principle of state self-determination (which further did not rest on the imprimatur of religious authorities.) As it faced the multiple challenges of the the rise of the autonomous state, increased religious competition, and massive warfare that was fought in the name of religion, the Church responded with greater attention to, and emphasis on, secular politics and the state.

These changes in the relative emphases of Church pronouncements challenges the "naïve doctrine hypothesis," which posits that the Church holds a set of views on church and state that reflect its core doctrinal teachings inherited from the foundational Church, and that these have not changed. If this were correct, we would expect to see little change in attention paid to state and social policy. Instead, we see both the increase in mentions of politics in over the 11-13th centuries, and a marked surge in discussions of the state and social policy after the 16th century. While we do not examine the *substance* of the documents in this iteration of the paper, the increase in relative proportions itself suggests that the Church hierarchy was paying more attention, and reasserting Church teachings, regarding politics and the state over time.

In two final notes, we observe differences both across popes and across languages. Within the groups of documents written by each Pope, we also see distinct emphases: for

example, note the steady decrease in emphasis on Liturgy and Devotions from the time of Pius IX (Pope from 1846 to 1878) to the modern times. Benedict XVI (2005-2013) actually issued relatively more declarations on the topic of Bioethics and Abortion, as well as Sex and Family, than Pope John Paul II (1978-2005), who was long seen as the most conservative and outspoken Pope on the topic. These patterns may be artefacts of the sample we have available for analysis: with a greater corpus, we may be able to trace such patterns more definitively.

Finally, we see a trend in the topic proportions by language. Documents about the church hierarchy, internal organization, and doctrine are disproportionately written in Latin, while documents about states and social policy are more often translated into the non-liturgical languages represented in our data. This may reflect a distinction between audiences: documents involving states and social policies are communicated directly to the laity, as they are more relevant to the daily lives of church members, while doctrinal pronouncements or documents relevant to church institutions are directed toward a different audience, namely the church leadership. But since Latin was the Church's *lingua franca* well into the 20th century, this may also reflect contemporary decisions made internally by the Church to retroactively translate its most important documents.

Conclusion

We conclude briefly by noting that these findings are simply the first step. A bigger corpus that incorporates a greater variety of documents, and which has greater coverage of the earlier period under consideration, may very well change these findings. Nonetheless, the patterns we observe are consistent with the notion that that the Roman Catholic Church increasingly turned its attention to the state over the course of the medieval period, and began to focus on it extensively in the 16th century. This new focus required making new conceptual distinctions, and responding not only to religious competition, but to the rise of an ever more powerful, and ever more autonomous, new secular state.

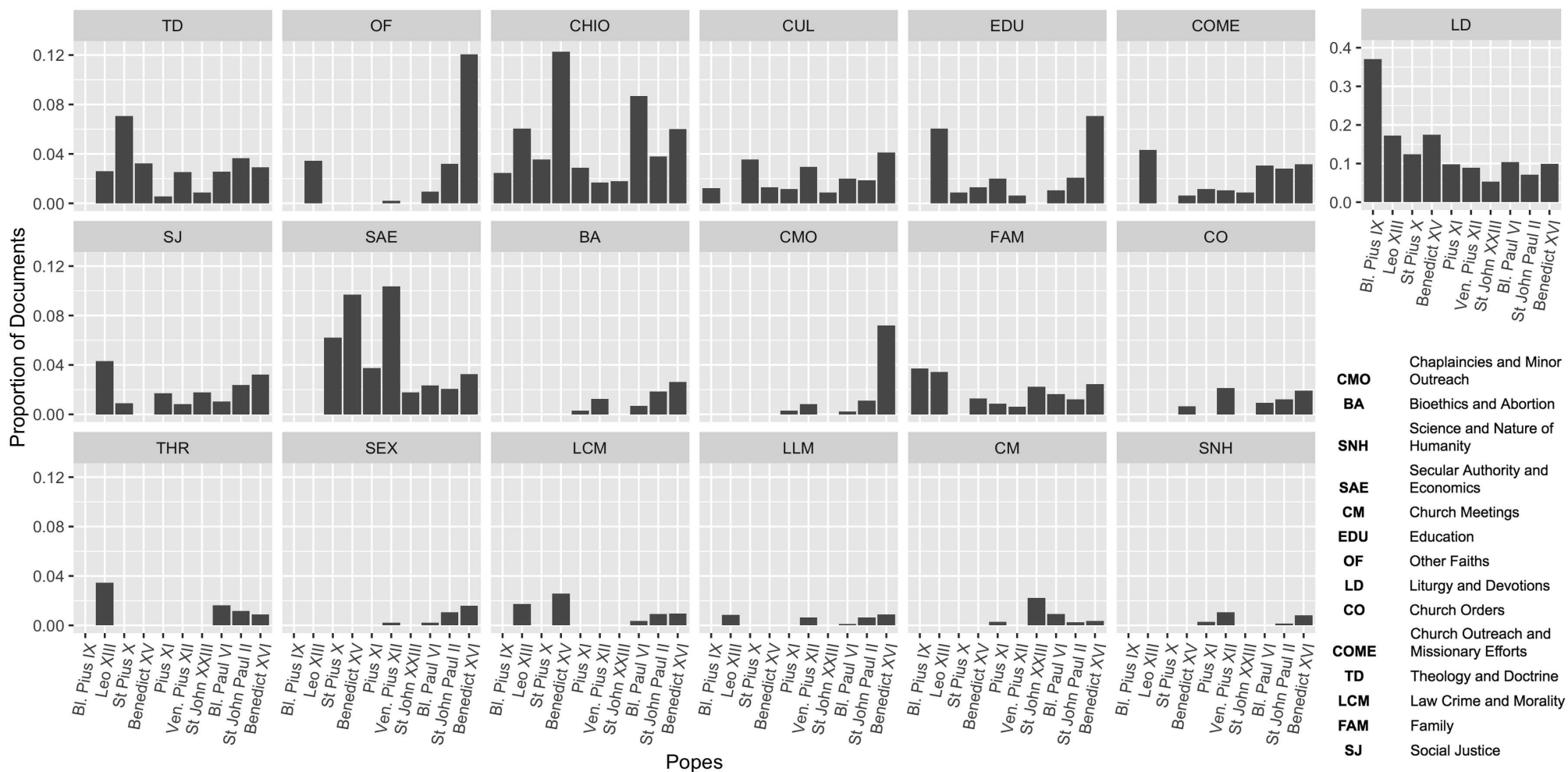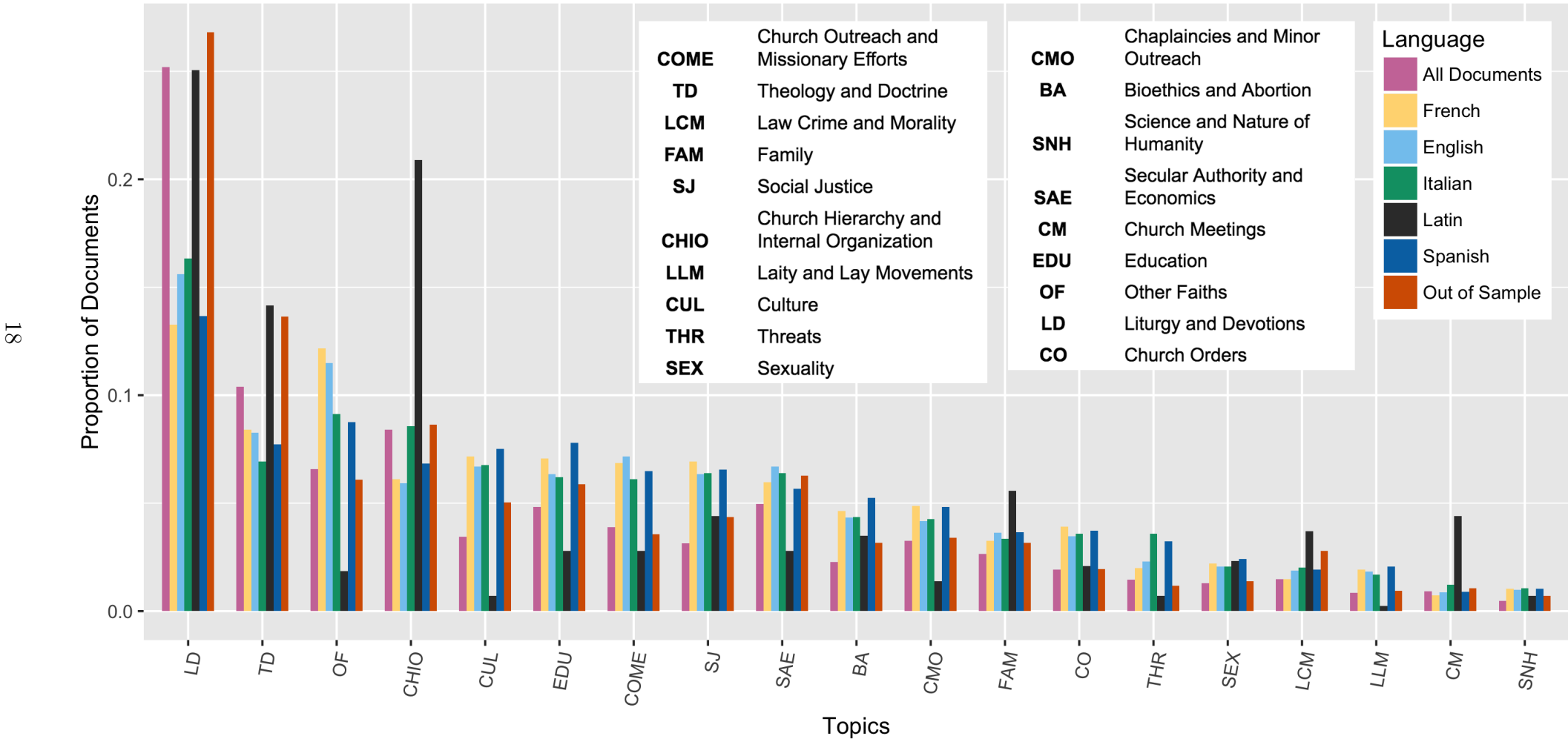Figure 4: Document Proportions by Pope

Figure 5: Document Proportions by Language and Subgroup

COME	Church Outreach and Missionary Efforts	**CMO**	Chaplaincies and Minor Outreach
TD	Theology and Doctrine	**BA**	Bioethics and Abortion
LCM	Law Crime and Morality	**SNH**	Science and Nature of Humanity
FAM	Family		
SJ	Social Justice	**SAE**	Secular Authority and Economics
CHIO	Church Hierarchy and Internal Organization	**CM**	Church Meetings
LLM	Laity and Lay Movements	**EDU**	Education
CUL	Culture	**OF**	Other Faiths
THR	Threats	**LD**	Liturgy and Devotions
SEX	Sexuality	**CO**	Church Orders

Appendix

Equations for Evaluation Metrics

$$\text{Recall} = \frac{TP}{TP+FN} \tag{5}$$

$$\text{Recall}_{\text{micro}} = \frac{(TP_1+...+TP_k)}{(TP_1+...+TP_2+FN_1+...+FN_k)} \tag{6}$$

$$\text{Recall}_{\text{macro}} = (\text{recall}_1 + ... + \text{recall}_k)/k \tag{7}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{8}$$

$$\text{Precision}_{\text{micro}} = \frac{(TP_1+...+TP_k)}{(TP_1+...+TP_k+FP_1+...+FP_k)} \tag{9}$$

$$\text{Precision}_{\text{macro}} = (\text{precision}_1 + ... + \text{precision}_k)/k \tag{10}$$

$$\text{F1} = 2 \cdot \frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}} \tag{11}$$

$$\text{F1}_{\text{weighted}} = \frac{(1+\beta^2)\cdot\text{recall}\cdot\text{precision}}{\text{recall}+\beta^2\cdot\text{precision}} \tag{12}$$

References

Armstrong, K. (2014, September). *The myth of religious violence.* Retrieved from https://www.theguardian.com/world/2014/sep/25/-sp-karen-armstrong-religious-violence-myth-secular

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(Feb), 281–305.

Blaydes, L., Grimmer, J., & McQueen, A. (2013). Mirrors for princes and sultans: Advice on the art of governance in the medieval christian and islamic worlds.

Cavanaugh, W. T. (2009). *The myth of religious violence: Secular ideology and the roots of modern conflict.* OUP USA.

Chaves, M., & Cann, D. E. (1992). Regulation, pluralism, and religious market structure explaining religion's vitality. *Rationality and society*, *4*(3), 272–290.

Chwe, M. S.-Y. (1994). Farsighted coalitional stability. *Journal of Economic theory*, *63*(2), 299–325.

Chwe, M. S.-Y. (2013). *Rational ritual: Culture, coordination, and common knowledge.* Princeton University Press.

Clark, J. N. (2010). Religion and reconciliation in bosnia & herzegovina: Are religious actors doing enough? *Europe-Asia Studies*, *62*(4), 671–694.

Esposito, J. L., & Mogahed, D. (2007). *Who speaks for islam?: What a billion muslims really think.* Simon and Schuster.

Esposito, R. (2010). Communitas: the origin and destiny of community.

Finke, R., & Stark, R. (1992). The churching of america 1776-1990. *New Brunswick*, *347*.

Fish, S. (2011). A letter concerning toleration. *The Pretenses of Loyalty: Locke, Liberal Theory, and American Political Theology*, 103.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer series in statistics Springer, Berlin.

Furnas, G. W., Deerwester, S., Dumais, S. T., Landauer, T. K., Harshman, R. A., Streeter, L. A., & Lochbaum, K. E. (1988). Information retrieval using a singular value decomposition model of latent semantic structure. In *Proceedings of the 11th annual international acm sigir conference on research and development in information retrieval* (pp. 465–480).

Gill, A. (2001). Religion and comparative politics. *Annual Review of Political Science*, *4*(1), 117–138.

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.

Iannaccone, L. R. (1998). Introduction to the economics of religion. *Journal of economic literature*, *36*(3), 1465–1495.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing*. Pearson.

Kahl, B. (2014). The bible, the economy, and the poor. *Journal of Religion & Society*, *133*, 10.

Kahl, S. (2005). The religious roots of modern poverty policy: Catholic, lutheran, and reformed protestant traditions compared. *Archives Européennes de Sociologie/European Journal of Sociology/Europäisches Archiv für Soziologie*, 91–126.

King James Bible. (n.d.). *King james bible*. Project Gutenberg.

Morgan, D. (2009). The look of sympathy: Religion, visual culture, and the social life of feeling. *Material Religion*, *5*(2), 132–154.

Perica, V. (2002). *Balkan idols: Religion and nationalism in yugoslav states*. Oxford University Press on Demand.

Smith, R. (2007). An overview of the tesseract ocr engine. In *Proceedings of the ninth international conference on document analysis and recognition - volume 02* (pp. 629–633). Washington, DC, USA: IEEE Computer Society. Retrieved from http://dl.acm.org/citation.cfm?id=1304596.1304846

Strayer, J. R. (1970). *On the medieval origins of the modern state*. Princeton University Press.

Woodberry, R. D. (2011). Religion and the spread of human capital and political institutions.