Leadership and Economic Growth: a Text Analytics Approach

Salfo Bikienga

Job Market Paper

October 25, 2016

Abstract

*In recent years, a growing number of economists have come to recognize the importance of political leadership in promoting economic performance. However, without an agreed upon measure of leadership, formally demonstrating and testing this relationship remains elusive. This paper proposes identifying economic leadership by measuring the consistency with which leaders talk about economic issues. We employ a text analytics approach–Topic Modeling–to studying leaders' discourses, and measure the relationship between these discourses and economic growth. Specifically, using the Latent Dirichlet Allocation (LDA) algorithm, we identified the topical content of U.S governors' state of the state speeches from 2001 to 2013, constructed a consistency measure over these topics, and studied the relationship between the consistency of these topical content and the states' real GDP growth. We find that the consistency with which governors address economic issues is strongly associated with economic growth. (**JEL C40, H70, O40**)*

1 Introduction

In recent years, the importance of political leadership in promoting economic growth and performance has gained increasing recognition in the economics profession (Brady and Spence, 2010). However, formally demonstrating and testing the relationship between leadership and growth remains elusive. Many economists remain skeptical of this seemly ephemeral concept, preferring to examine concrete policy actions. This paper demonstrates that text analytics is a viable approach for identifying leadership and testing its role in promoting economic growth. We identify economic leadership through textual analysis of their public speeches and then test whether leader's focus on relevant economic topics is related to subsequent economic performance. In particular, we analyze the “state of the state” speeches of U.S governors from 2001 to 2013. A governor is economically successful if his/her tenure was marked by a positive economic performance (e.g. average positive real GDP growth deviation from the U.S average). We find that governors who persistently talk about the economy significantly perform better than their counterparts.

This paper provides evidence that leadership matters and that leadership can be inferred from public speeches. Indeed, the role of leadership in the success of organizations is widely recognized in the management literature (Lieberson and O'Connor (1972); Thomas (1988); Jing and Gayle (2008)). However, with a few exceptions (Brady and Spence (2010); Jones and Olken (2005)), economists have typically remained skeptical of the role of leadership. To the extent that leader's preferences and priorities shape both, direct policy actions and economic institutions, these priorities can be critical for economic growth, (Byman and Pollack (2001)). The leader's priorities may be taken as proxies for the myriad actions taken by the leaders to encourage growth through appointments, setting the tone for governmental agencies and promoting legislation. These positive actions are difficult to measure directly and hence we seek to measure them indirectly through analysis of the leader's priorities, as expressed in his/her speeches.

Textual analysis has become a standard tool in the social sciences, particularly in the political science literature (LAVER et al. (2003); Quinn et al. (2010); Wilkerson et al. (2015)) , but is relatively new to the economics literature (Zubin et al. (2014); Baker et al. (2013); Alexopoulos and Cohen (2015)).

In this paper, we focus on the relationship between the priorities of U.S governors and state economic growth. Compared to national leaders in developing countries, U.S. governors have far less power to promote favorable economic policies. The main advantages of using U.S. governors is that most present a formal, annual, "state of the state" speech and the overwhelming majority of these speeches are readily available. We thus avoid many of language, cultural and political differences that make international comparisons problematic.

The remainder of the paper is organized as follows: section 2 reviews the debates about the role of political leaders for societal progress, and presents a framework informing on the usefulness of text analytics for studying the influence of political leaders. Section 3 explains the statistical method used in this paper. Section 4 describes the data, and section 5 presents the results. Section 6 concludes the paper.

2 Do political Leaders matter?

2.1 Leadership traits

Hermann et al. (2001) proposes a framework for analyzing political leaders. The framework identifies four types of leaders, in a two-dimensional scale: the crusader, the strategist, the pragmatist, and the opportunist. The two dimensions are: a) the way leaders challenge constraints, and b) the way they are open to new information. The goal-driven leaders (crusaders and strategists) interpret environmental constraints through a lens that is structured by their beliefs, motives, and passions. They see constraints as obstacles in their way, and must be overcome. Goals are to be achieved by all available means. Policy priorities are clearly defined and collaborators are chosen on the basis of their general belief and support of what the leader perceives to be best for all concerned parties. Information is filtered in accordance with the government's policies, rather than their objectiveness. "We know what we want, and we only need information telling us how to get it". They focus on achieving their goals; and because of their focus, they are more likely to be consistent in what they say. On the contrary, the more responsive leaders (pragmatists and opportunists) see life as "a theater where there are many roles to be played" and they avoid taking action unless the option chosen is supported by their constituencies. Principles and goals are sacrificed for the sake of consensus building. For such leaders, constraints set the parameters for action. Whereas a leader with agendas seeks information that reinforces their beliefs, the responsive leader is interested in what is possible under the current circumstances; and, as the saying goes, "runs an idea up the flagpole to see who salutes it." They are like chameleons, which change their stance according to the situation. Inaction is preferred to an action that may discontent constituents. They cannot be consistent in what they say, since they respond to circumstances. In this paper, we will identify the governors' consistency (during their tenures) over a constructed list of topics, and analyze the relationship between the consistency measure and economic growth.

2.2 Text analytics and leaders' discourses

Consistency in what one says is assessed through one's discourse. Text analytics allows a systematic study of discourses, and it can be used to capture leaders' expressed priorities. In fact, the desire to understand, and predict the behavior of political leaders has compelled political scientists to apply statistical methods on leaders' discourses to characterize their leadership style (Hermann et al. (2001)), or to determine their expressed agenda (Grimmer (2010)). These methods, also referred as content analysis, are intended to extract leaders' motives at distance (as opposed to surveying leaders). These methods are particularly useful for political leadership studies for several reasons: we cannot reach political leaders to administer surveys; we do not have clearly defined

Table 1: Words count in documents (Document Term Matrix)

	and	but	economics	hate	hates	he	i	love	math
document1	0	1	1	1	0	0	2	1	1
document2	1	0	1	0	2	2	0	0	1

leadership variables and data on these variables. The "one kind of data from political leaders that is produced and preserved in abundance" is their words (Winter (2005)). Political leaders communicate their agenda, mobilize followers, and research suggests that their public statements reflect what they want, and what they are pledging to be (Hermann (2008)). Thus, text analytics gives us a viable means for quantifying leaders' expressed priorities, which then permits research to explore the relationships of these priorities to economic growth.

3 Methodology

3.1 Text Analytics

This paper uses text analytics methodology to infer topics covered in U.S. governors' speeches, and analyze the correlation between these topics and the real GDP growth. Text analytics aims to extract useful information from text documents and do so in a formal, automatic manner. Text analytics consists of the application of statistical methodologies to textual sources (Solka (2008)). The idea of text analytics is to apply statistical methods, designed to analyze numbers, to words. One of the main tasks for text analytics is to transform unstructured texts into numerical data. As an example of how to convert words into a spreadsheet, assume the following two sentences constitute two documents (document 1 and 2):

[document1] "i love economics, but i hate math"

[document2] "he hates economics, and he hates math"

A spreadsheet of words count can be created as shown in table 1:

Once, the data table is created, the remaining of the analysis is just an application of several of traditional and modern statistical tools. By modern statistical tools, we refer to statistical learning or machine learning tools (Varian (2014)). Text analytics is widely used in social sciences, especially in political science to analyze political speeches and legislation (LAVER et al. (2003); Quinn et al. (2010); Wilkerson et al. (2015)). A few examples of the use of text analytics can be found in economics too. Zubin et al. (2014) shows that political ideology influences economic research in the U.S by using "observed political behavior of economists and the phrases from their academic articles" to construct predictors "of political ideology by article, economists, school and journal." Baker et al. (2013) proposes a policy uncertainty index, based "on the frequency of newspaper references to policy uncertainty and other indicators." Similarly, Alexopoulos and Cohen (2015) proposes general economic uncertainty, and policy uncertainty indicators, "based on textual analysis of information contained in The New York Times".

One challenge in text analytics is the dimensionality of the data; if words are the variables, and there are possibly thousands of different words for a given collection of documents, then the number of variables is unusual for traditional statistical tools (For example, OLS breaks down when $n \leq p$, n being the number of observations and p, the number of variables, or words). Consequently, we resort to the use of dimensionality reduction methods[1]. Text analytics is an extensive field, which ranges from key words finding and analysis (Romer and Romer, 2015), to inferring themes (or topics) from text documents. This paper uses topic modeling because it produces the relative importance (proportions) of topics covered in text documents. Thus, by using topic modeling, we can assess the relative importance of each topic for each governor over time.

3.2 Topic models

Topic modeling derives from Latent Semantic Allocation (LSA), which is a linguistics theory of meaning that uses linear algebra to collapse words in a collection of documents into clusters of words (Landauer et al. (2007), chap1&2). The clusters of words are meant to represent themes in the documents. LSA postulates that meaning stems from words co-occurrence regardless of syntax. Thus, by a matrix factorization, from a matrix of thousands of words (variables), it is possible to reduce it to a matrix of a few topics, a matrix of their relative importance represented by their eigenvalues, and a new words-matrix. The new words-matrix provides clues for naming the topics. This paper uses a Bayesian matrix factorization algorithm known as Latent Semantic Allocation (Blei et al. (2003)). A further exposition of topic modeling can be found in appendix A.1. For the sake of exposition, the matrix decomposition below (a reduced form) is given for illustration.

$$
\begin{array}{cc}
 & \begin{array}{cccc} w_1 & w_2 & \dots & w_V \end{array} \\
\begin{array}{c} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_D \end{array} &
\begin{pmatrix}
n_{1,1} & n_{1,2} & \dots & n_{1,V} \\
n_{2,1} & n_{2,2} & \dots & n_{2,V} \\
n_{3,1} & n_{3,2} & \dots & n_{3,V} \\
\vdots & \vdots & \vdots & \vdots \\
n_{D,1} & n_{D,2} & \dots & n_{D,V}
\end{pmatrix}
\end{array}
\approx
\begin{array}{cc}
 & \begin{array}{cc} t_1 & t_2 \end{array} \\
\begin{array}{c} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_D \end{array} &
\begin{pmatrix}
\theta_{1,1} & \theta_{1,2} \\
\theta_{2,1} & \theta_{2,2} \\
\theta_{3,1} & \theta_{3,2} \\
\vdots & \vdots \\
\theta_{D,1} & \theta_{D,2}
\end{pmatrix}
\end{array}
*
\begin{array}{cc}
 & \begin{array}{cccc} w_1 & w_2 & \dots & w_V \end{array} \\
\begin{array}{c} t_1 \\ t_2 \end{array} &
\begin{pmatrix}
\phi_{1,1} & \phi_{1,2} & \dots & \phi_{1,V} \\
\phi_{2,1} & \phi_{2,2} & \dots & \phi_{2,V}
\end{pmatrix}
\end{array}
$$

The first matrix (left side) is a matrix of word frequencies for a collection of D documents, and a

[1] Varian (2014) presents a quick survey of these methodologies, highlighting their usefulness for applied economics. Einav and Levin (2014) surveys the use of big data for applied economics.

list of V words. $n_{1,1}$ is the frequency of word w_1 in the document d_1. The second matrix (middle) is the matrix of topic distributions in the documents. For example, $\theta_{1,1}$ is the proportion of topic 1 (t_1) in document d_1. This matrix (the matrix of the θs) preserves almost all essential information about the documents that the words in the first matrix contain. For example, by using the two dimensions, t_1 and t_2 of this matrix, we can analyze document similarities with a simple scatterplot. The third matrix (right side) gives the words distributions in each topic. For example, $\phi_{1,1}$ gives the relative importance of word w_1 for the topic t_1. To name topic 1, we need to sort the first row of the third matrix by decreasing order. Ideally, the first few words will identify a recognizable concept. For example, if the first few words of topic t_1 are: education, college, tuition, teacher, we may conclude that topic t_1 is about education. And $\theta_{1,1}$ is the relative importance of education in document d_1. Intuitively, the goal of the LDA algorithm is to iteratively try different values for the θs and ϕs untill their joint product is highest. Thus, the algorithm searches for the θs and ϕs for which the likelihood of observing the given collection of documents is highest. The usefulness of topic modeling for the current paper is its ability to automatically provide the topics distributions $\theta_{d,k}$, d being a governor's speech, and k being a topic. Topic modeling informs on the topics, and their relative importance in every leader's speech. Knowing the relative importance of each topic in each leader's speech, and how the importance of topics changes over time can inform on the priorities of a leader.

4 Data

4.1 Text data

We choose to use the state of the state speeches of U.S governors from 2001 to 2013 for two reasons. Most of them can be accessed online, and they are given at a specific time of the year. Therefore, they can be used to compare governors. Most of the speeches (500 of the 598 speeches) were automatically collected (scrapped) from the state of the state website. Most of the remaining speeches were collected from The Pew Charitable Trusts website. A few of the speeches were collected from the governors' websites. Once the speeches are gathered, the next step consists of preprocessing the data:

- Convert all words to lower cases, to avoid two identical words being considered different because one of them uses a capital letter;
- Remove stop words, which are words such as a, to, for, and, ...; they do not add content to texts.
- Strip white spaces, which is to remove unnecessary spaces and tabs in a text;

- Drop words of less than four characters; most of them do not add content to texts.
- Remove punctuation and numbers;
- Take words stems, i.e. take the roots of words to avoid, for example, economy, economics, economical to be considered as three different words.

Once the preprocessing is done, we create the Document-Term-Matrix (DTM), that is, our data matrix. The text documents (or speeches) are now converted to a spreadsheet with words counts in the cells. The DTM is then fed into the LDA algorithm to get the θ and ϕ matrices. The θs are used for the remainder of the analysis. The ϕs are used to interpret the topics.

4.2 Economic data

We use the state average real GDP growth rate deviation from the U.S growth rate (period 2001 to 2014) as our dependent variable in the analysis. The data were downloaded from the Bureau of Economic Analysis (BEA) website.
The real GDP percent growth variable is measured at the end of the year; the speeches are delivered at the beginning of the year, usually in January. Thus, the speech and the growth rate can be seen as one year apart. We may speculate that if a speech informs on what a leader intends to do, the effects of the leader's actions are only visible after several months. We further use two-lead period in the analysis.

4.3 Ideology data

The government ideology data, compiled and maintained by Berry et al. (1998) was also collected to study the interaction effect of the state's government ideology and the economic agenda variable constructed from the text data. The data set spans from 1960 to 2014 and provide an annual index of the ideology of each state government. The index ranges from 0 to 100, with 0 representing the most conservative government (i.e. the legislative and executive power is completely controlled by Republicans), and 100 the most liberal government. This paper converts the continuous index into three categories, under the assumption that an ideological index of 60 or above confers the Democrat governor a great political power. Similarly, an index of 40 or less confers the republican governor a great power. An index between 40 and 60 confers the governor a moderate power.

Table 2 shows that most state governments are ideologically controlled by either democrats or republicans (77 of 102 cases).

Table 2: Counts of Ideological Control of State Governments

Power_D	Power_N	Power_R
41	25	36

4.4 Variables construction

One goal of this paper is to show that text analytics is a viable tool for studying political leadership and economic growth. To do so, we study the correlation between governors' consistency over certain topics (which is assumed to be a proxy for governors' priorities) and economic growth. The inverse of the coefficient of variation (CV) of a topic is used as our consistency measure ($C_{i,j}$). Formally, the consistency over a topic j is given by:

$$C_{i,j} = \frac{\bar{X_{i,j}}}{s_{i,j}},$$

where $\bar{X_{i,j}}$ is the average proportion of topic j in the speeches of a given governor i, and $s_{i,j}$ is the standard deviation of that topic for that governor. Intuitively, a governor who talks profusely and consistently about a topic should have a high average and low variance for that topic. We limit the data to governors with at least three speeches; the choice of three speeches is to assure we have enough observations to compute meaningful means and variances. 102 governors satisfy this condition (i.e. 102 observations). The consistency measures are the independent variables, and there are 5 topics. Based on a ten folds cross validation approach, 5 topics is the optimum number of topics for our data (see appendix B for the details). Next, we compute the state' real GDP growth rate deviation from the U.S, followed by their averages by governor.

$$\bar{g}_{governor_i} = \frac{1}{tenure} \sum_{l=1}^{tenure} (g_{state_{i,l}} - g_{US_i}),$$

where $\bar{g}_{governor_i}$ is the average state growth (g_{state}) deviation from the US growth rate (g_{US}). The averages are computed by governor's tenure.

Table 3 presents the summary of the final variables. Topic.1 and Topic.5 have a few outliers. Removing them from the analysis does not change the main results.

5 Results

In sum, the text data is converted into a matrix of words counts, which is then used to generate clusters of words that represent the topics (5 topics). Each topic is then converted into a consistency

Table 3: Data Summary Table

Statistic	N	Mean	St. Dev.	Min	Max
Topic.1	102	5.561	5.611	1.081	43.573
Topic.2	102	3.560	2.802	1.235	17.029
Topic.3	102	4.142	2.558	1.426	15.059
Topic.4	102	3.814	2.787	1.004	18.338
Topic.5	102	3.961	3.845	1.165	34.427
gdp_r_perc_changeDv1	102	−0.137	1.166	−3.233	4.733
gdp_r_perc_changeDv2	102	−0.126	1.145	−2.800	3.300

measure ($C_{i,j} = \frac{\bar{X}_{i,j}}{s_{i,j}}$). Each governor performance is capture by $\bar{g}_{governor_i}$. The final regression equation is:

$$\bar{g}_{governor_i} = \beta_0 + \sum_{j=1}^{5} \beta_j C_{i,j} + \varepsilon_i,$$

$C_{i,j}$ being the topic j consistency measure for governor i.

5.1 One period lead growth variable

The dependent variables used are the one and two-period leads of the state average real GDP growth rate, deviation from the U.S average real GDP growth rate. The following chart illustrates the one and two-period leads idea. For instance, assuming a one term governor with speeches from beginning 2001 to the beginning 2004, his/her agenda (i.e. the governor's consistency measure over a topic) is matched with the average growth rate of the end of 2001 to the end of 2004 for the one period lead dependent variable, and the average growth rate of the end of 2002 to the end of 2005 for the two-period lead dependent variable.

		2 period lead
	1 period lead	$g_{i,5}$
Speech	$g_{i,4}$	$g_{i,4}$
4th speech	$g_{i,3}$	$g_{i,3}$
3rd speech	$g_{i,2}$	$g_{i,2}$
2nd speech	$g_{i,1}$	
1st speech		

5.1.1 Economic agenda and economic growth

Table 4, column (1) shows the result of the OLS regression of the five topics on the one period lead of the state real GDP average deviation from US growth ($\bar{g}_{governor_i}$). Only one topic (topic.4) appears significant. We further use the LASSO (Least Absolute Shrinkage and Selection Operator) method to drop the least relevant topics. The LASSO is a constrained OLS. The constraint is such that it sets the parameters of the non-relevant exogenous variables to zero, yielding a sparse model. It is one of the most used and reliable variable selection methods. LASSO can be seen as an alternative to the stepwise regression method; however it is more principled than the stepwise approach, as it is a model optimization based method.

Of the 5 topics, the LASSO regression picked 2 topics, that is, we can safely ignore three topics in our final regression model. By fitting an OLS on the selected 2 topics, the results are as shown in the second column of Table 4. Note that dropping the three topics from the model does not change the results, confirming that these three topics are not needed in the regression model.

Topic.4 is a variable of interest because it is about governors' economic agenda as stated in their speeches, which we suspected might affect the state growth rate. The reason Topic.4 is called governor's economic agenda is explained in the next section. Before then, column 3 of Table 4 shows the results for a different definition of the consistency variable (the standard deviation of the topic). The idea is that too much variation on a particular topic is a sign of lack of focus (or lack of agenda). The Topic.4 coefficient is negative and statistically significant suggesting that the lack of focus on the governor economic agenda is associated with negative economic growth.

A quick graphical look at the data (scatter plot of growth variable and Topic.4) shows two outliers values for Topic.4. To check whether the significance of Topic.4 is affected by the presence of these outliers, the regression model is run (1) using the log of the topics (column 2 of Table 5), (2) dropping the two outliers from the data and running the regular model (column 3). Topic.4 remains statistically significant under these two specifications.

It appears that Topic.4 is statistically significant however we look at its relationship with the growth variable. This topic is about economic development, and is positively related to the growth variable. It can be speculated that a governor with an economic development agenda is more likely to achieve a positive economic growth. Why Topic.4 can be understood as the economic agenda topic?

5.1.2 Interpreting Topic.4 as Economic Agenda Topic

Traditionally, topics from topic modeling are interpreted by looking at the most frequent words that constitute a topic, usually the first 30 words. We rely on three different approaches to interpreting the topics in this paper.

First, by looking at the first 40 words (Appendix C, table 12), which are ranked by their rel-

Table 4: OLS with 5 topics (1), with 2 topics picked by LASSO (2), and using topics' standard deviation (3)

	Dependent variable:		
	One period lead of state average real GDP growth (deviation from US)		
	(1)	(2)	(3)
Topic.1	−0.006		
	(0.022)		
Topic.2	0.010		
	(0.041)		
Topic.3	−0.059	−0.057	−2.483
	(0.045)	(0.043)	(3.801)
Topic.4	0.134***	0.135***	−8.375**
	(0.040)	(0.040)	(4.090)
Topic.5	0.015		
	(0.032)		
Constant	−0.460	−0.418	0.524
	(0.314)	(0.255)	(0.361)
Observations	102	102	102
R^2	0.120	0.118	0.046
Adjusted R^2	0.075	0.100	0.026
Residual Std. Error	1.122 (df = 96)	1.107 (df = 99)	1.151 (df = 99)
F Statistic	2.627** (df = 5; 96)	6.603*** (df = 2; 99)	2.371* (df = 2; 99)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: OLS with all observations (1), taking the log of Topic.4 (2), removing the two outliers (3)

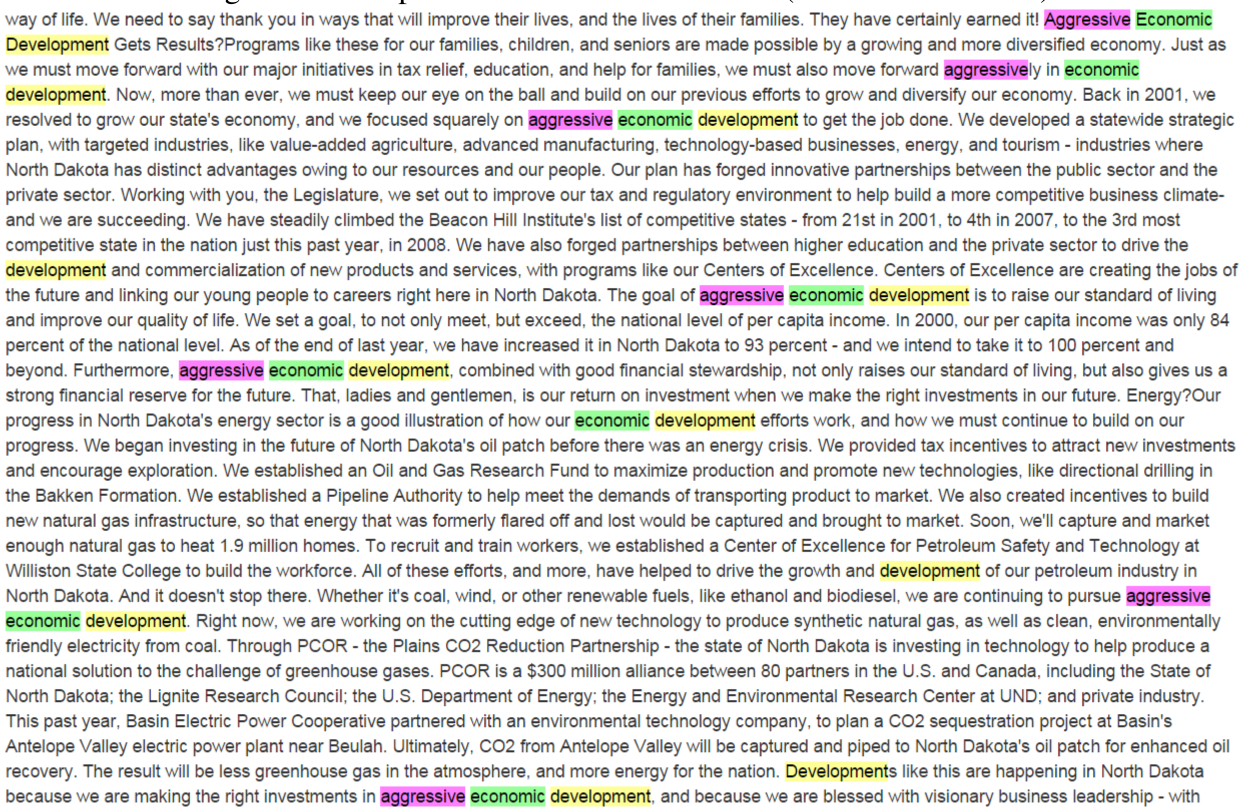
	Dependent variable:		
	One period lead of state average real GDP growth (deviation from US)		
	(1)	(2)	(3)
Topic.3	−0.057	−0.241	−0.070*
	(0.043)	(0.221)	(0.041)
Topic.4	0.135***	0.602***	0.087*
	(0.040)	(0.207)	(0.052)
Constant	−0.418	−0.531	−0.209
	(0.255)	(0.378)	(0.273)
Observations	102	102	100
R^2	0.118	0.086	0.057
Adjusted R^2	0.100	0.067	0.037
Residual Std. Error	1.107 (df = 99)	1.127 (df = 99)	1.049 (df = 97)
F Statistic	6.603*** (df = 2; 99)	4.631** (df = 2; 99)	2.928* (df = 2; 97)
Note:			*p<0.1; **p<0.05; ***p<0.01

ative importance for the topic, it can be seen that economically related words are highly ranked for Topic.4. Words such as business, work, create, energy, develop, company, invest, industry, company... are highly ranked.

Second, we constructed a web application aimed at highlighting topic's keywords within the speeches that are highest on a given topic (Figure 2). Again, a speech is a distribution over topics. Some speeches are very high on a particular topic (for instance, over 60% of the speech is about a particular topic), while others are low on that particular topic. Thus, for a given topic, by identifying the speeches with the highest proportions of that topic and highlighting the topic's keywords in those speeches, we may be able to fully identify the topic. The top five documents for Topic.4 are filled with "Economic development" phrases. One such phrase is: " aggressive economic development". It should be noted that the top five documents are from a single state (North Dakota). Figure 1 is an excerpt from the 2009 state of the state speech of the governor of North Dakota. This excerpt shows a clear delineation of economic agenda.

It is true that North Dakota had an oil boom that started in 2006, but governor John Hoeven economic agenda was prominent even in his earlier speeches. The word "develop" and its variants were used more than ten times in his 2001 state of the state speech. The phrase "New economy" was used 9 times in the same speech. Further, Topic.4 has been high in all of his speeches, ranging

Figure 1: Excerpt of the state of state address (North Dakota 2009)



way of life. We need to say thank you in ways that will improve their lives, and the lives of their families. They have certainly earned it! Aggressive Economic Development Gets Results?Programs like these for our families, children, and seniors are made possible by a growing and more diversified economy. Just as we must move forward with our major initiatives in tax relief, education, and help for families, we must also move forward aggressively in economic development. Now, more than ever, we must keep our eye on the ball and build on our previous efforts to grow and diversify our economy. Back in 2001, we resolved to grow our state's economy, and we focused squarely on aggressive economic development to get the job done. We developed a statewide strategic plan, with targeted industries, like value-added agriculture, advanced manufacturing, technology-based businesses, energy, and tourism - industries where North Dakota has distinct advantages owing to our resources and our people. Our plan has forged innovative partnerships between the public sector and the private sector. Working with you, the Legislature, we set out to improve our tax and regulatory environment to help build a more competitive business climate- and we are succeeding. We have steadily climbed the Beacon Hill Institute's list of competitive states - from 21st in 2001, to 4th in 2007, to the 3rd most competitive state in the nation just this past year, in 2008. We have also forged partnerships between higher education and the private sector to drive the development and commercialization of new products and services, with programs like our Centers of Excellence. Centers of Excellence are creating the jobs of the future and linking our young people to careers right here in North Dakota. The goal of aggressive economic development is to raise our standard of living and improve our quality of life. We set a goal, to not only meet, but exceed, the national level of per capita income. In 2000, our per capita income was only 84 percent of the national level. As of the end of last year, we have increased it in North Dakota to 93 percent - and we intend to take it to 100 percent and beyond. Furthermore, aggressive economic development, combined with good financial stewardship, not only raises our standard of living, but also gives us a strong financial reserve for the future. That, ladies and gentlemen, is our return on investment when we make the right investments in our future. Energy?Our progress in North Dakota's energy sector is a good illustration of how our economic development efforts work, and how we must continue to build on our progress. We began investing in the future of North Dakota's oil patch before there was an energy crisis. We provided tax incentives to attract new investments and encourage exploration. We established an Oil and Gas Research Fund to maximize production and promote new technologies, like directional drilling in the Bakken Formation. We established a Pipeline Authority to help meet the demands of transporting product to market. We also created incentives to build new natural gas infrastructure, so that energy that was formerly flared off and lost would be captured and brought to market. Soon, we'll capture and market enough natural gas to heat 1.9 million homes. To recruit and train workers, we established a Center of Excellence for Petroleum Safety and Technology at Williston State College to build the workforce. All of these efforts, and more, have helped to drive the growth and development of our petroleum industry in North Dakota. And it doesn't stop there. Whether it's coal, wind, or other renewable fuels, like ethanol and biodiesel, we are continuing to pursue aggressive economic development. Right now, we are working on the cutting edge of new technology to produce synthetic natural gas, as well as clean, environmentally friendly electricity from coal. Through PCOR - the Plains CO2 Reduction Partnership - the state of North Dakota is investing in technology to help produce a national solution to the challenge of greenhouse gases. PCOR is a $300 million alliance between 80 partners in the U.S. and Canada, including the State of North Dakota; the Lignite Research Council; the U.S. Department of Energy; the Energy and Environmental Research Center at UND; and private industry. This past year, Basin Electric Power Cooperative partnered with an environmental technology company, to plan a CO2 sequestration project at Basin's Antelope Valley electric power plant near Beulah. Ultimately, CO2 from Antelope Valley will be captured and piped to North Dakota's oil patch for enhanced oil recovery. The result will be less greenhouse gas in the atmosphere, and more energy for the nation. Developments like this are happening in North Dakota because we are making the right investments in aggressive economic development, and because we are blessed with visionary business leadership - with

from about 49% to about 69%; that is to say, economic development has been his main focus during his tenure.

Last, we introduced external speeches (speeches we know should be high on the economic topic) to our collection of governors' speeches. We expect the algorithm to assign high proportions to the economic topic (Topic.4) in these speeches. Our interpretation of Topic.4 is validated if the topic modeling algorithm assigns high proportion to the economic topic in these speeches. Ten economic policy speeches by four different politicians (George W. Bush (3), Obama (3), Romney (1), McCain (1), Jeb Bush (1), Hilary Clinton (1)) were added to our collection of speeches[2]. All these speeches are fairly high on Topic.4 (Table 6), supporting the interpretation that Topic.4 is indeed about economic agenda. It suffices to look at the totals rows of table 6 to see that Topic.4 and Topic.1 are high for these ten speeches. The totals of the remaining topics are fairly low.

Topic.1 is what we will call a residual topic, that is, a topic made up of mostly contentless words. Topic.2 is about education, and Topic.5 is about the state budget. Topic.3 is a mix of education, budget, economy with a reform tone. A disproportionate space was given in this paper for the interpretation of Topic.4 because it is the one which is statistically significant however we look at the relationship between the growth variable and the consistency measure of the topics.

[2]Web links to the ten speeches can be found in Appendix D

Table 6: Topics distributions in documents labeled by their authors as speeches on the economy

	Topic.1	Topic.2	Topic.3	Topic.4	Topic.5
B_Obama_2010	0.272	0.079	0.203	0.376	0.071
B_Obama_2012	0.429	0.044	0.148	0.345	0.035
B_Obama_2013	0.422	0.121	0.104	0.309	0.044
GW_Bush_2003	0.297	0.023	0.221	0.367	0.092
GW_Bush_2006	0.514	0.103	0.049	0.268	0.066
GW_Bush_2008	0.202	0.040	0.217	0.329	0.212
H_Clinton_2015	0.349	0.081	0.170	0.356	0.043
Jeb_Bush_2015	0.395	0.062	0.095	0.364	0.085
J_McCain_2008	0.355	0.035	0.130	0.331	0.149
M_Romney_2012	0.384	0.068	0.287	0.221	0.041
Totals	3.617	0.655	1.624	3.266	0.837

It should be noted that the topics are not defined ex-ante. Before the analysis, we had no knowledge of the topics covered in the speeches, and the number of topics that should be considered. The analysis suggested we collapse the words into only five topics for analyzing the growth variable, and it turned out that the topic that mostly matters for economic growth is the economic development topic. This strongly suggests that the governor's economic agenda matters for economic growth.

We further perform the analysis by interacting the economic agenda variable with the governor's political affiliation and find that both, democrats and republicans economic agenda correlate positively with economic growth (Table 7); however the correlation is stronger for republicans than for democrats. Column 1, of table 7 shows the results when no particular treatment of the outliers is considered, and column 2 shows the results when we take the log of the topics as a way to temper the effects of the outliers on the results.

5.1.3 Economic Agenda, Political Power and Economic growth

Does ideological control of the government influences the effect of governor's economic agenda? It can be argued that the governor's agenda will be difficult to implement without an ideological control of the government. For instance, a Democrat governor operating within a context in which Democrats control the legislative branch of the government has more leeway in implementing his/her agenda. Table 8 presents an analysis of the economic agenda effect on economic growth, by political party interacted with ideological control of the government.

The results suggest that the economic agenda of a Republican governor whose political party controls the government is positively associated with economic growth. The result is not significant

Table 7: Interaction with Governor's Political Affiliation. OLS without consideration of outliers (1), OLS using the log of Topic.4 (2)

	Dependent variable:	
	One period lead of state average real GDP growth(deviation from US)	
	(1)	(2)
Topic.3	-0.054	-0.212
	(0.044)	(0.226)
Topic.4:PartyD	0.094^{**}	0.519^{**}
	(0.045)	(0.222)
Topic.4:PartyR	0.177^{***}	0.697^{***}
	(0.046)	(0.229)
Constant	-0.426^{*}	-0.569
	(0.254)	(0.381)
Observations	101	101
R^2	0.143	0.093
Adjusted R^2	0.116	0.065
Residual Std. Error (df = 97)	1.099	1.130
F Statistic (df = 3; 97)	5.394^{***}	3.325^{**}
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 8: Interaction with Governor's Political Affiliation and Ideology. OLS without a special treatment of the outliers (1), OLS after removing the two outliers (2), OLS after taking the log of Topic.4 (3)

	Dependent variable:		
	One period lead of state average real GDP growth (deviation from US)		
	(1)	(2)	(3)
Topic.3	−0.040	−0.052	−0.091
	(0.042)	(0.042)	(0.224)
Topic.4:PowerPower_D:PartyD	0.057	0.026	0.340
	(0.045)	(0.061)	(0.225)
Topic.4:PowerPower_N:PartyD	0.247^{**}	0.198^{*}	0.916^{***}
	(0.100)	(0.102)	(0.345)
Topic.4:PowerPower_R:PartyD	0.232	0.189	0.836
	(0.226)	(0.222)	(0.725)
Topic.4:PowerPower_D:PartyR	0.116	0.044	0.526
	(0.225)	(0.224)	(0.674)
Topic.4:PowerPower_N:PartyR	−0.073	−0.124	−0.059
	(0.109)	(0.111)	(0.374)
Topic.4:PowerPower_R:PartyR	0.201^{***}	0.117^{**}	0.802^{***}
	(0.045)	(0.056)	(0.224)
Constant	-0.430^{*}	−0.204	-0.650^{*}
	(0.258)	(0.279)	(0.379)
Observations	101	99	101
R^2	0.241	0.143	0.187
Adjusted R^2	0.184	0.077	0.126
Residual Std. Error	1.056 (df = 93)	1.030 (df = 91)	1.093 (df = 93)
F Statistic	4.222^{***} (df = 7; 93)	2.163^{**} (df = 7; 91)	3.064^{***} (df = 7; 93)
Note:			*p<0.1; **p<0.05; ***p<0.01

for a Democrat governor whose political party controls the government, though the relationship remains positive. By ideological control, we mean at least 60% of the state legislative and executive power is held by one political party. The data suggests that Democrat governors' economic agenda matters when they do not have ideological control of the government. The results hold under different specifications (removing outliers values from the dataset (column 2), and using the log values of Topic.3 and Topic.4 column (3)). This analysis uses 101 observations instead of 102 governors because one governor is independent. The scatterplots (Figure 2) show the relationship between economic growth and economic agenda by governor political party, and the ideological dominance of the state government.

The scatterplots offer a visual prospect of the relationships shown in the regression table (Table 8).

5.2 Two-period lead growth variable

The analysis in section 5.1 shows that a governor's economic agenda matters for economic growth by using a one lead growth rate. Next, we perform the same analysis with a two-period lead growth variable. The result of the OLS regression using K = 5 topics is given in Table 9, column (1).
Again, Topic.4 seems to be strongly relevant for the two-period lead economic growth variable. We further use LASSO regression to drop the least relevant topics, and LASSO picked two topics of which the economic development topic (Topic.4), and the budget topic, Topic.5 are statistically significant (Table 9, column (2)). Column 3 of Table 9 shows the regression results when the two outliers values of Topic.4 are removed from the data set; and column 4 shows the results when we take the log of the dependent variable.
Table 9 shows that the economic development agenda is statistically significant and positive however we deal with the outliers values. This suggests that the significance of Topic.4 must be meaningful. Also, it can be said that the governor's economic agenda has a lasting effect on economic growth.

5.2.1 Economic Agenda, Political Power and Economic growth

We further the analysis by taking into consideration the governor political party and the ideological control of the government. The results are similar to the results performed using the one lead growth rate as the dependent variable. The results in Table 10 are for the regular model (column 1), removing the two outliers values of Topic.4 (column 2), and taking the log of the topics (column 3).

Again, as shown in section 5.1.3, the governor's economic agenda matter for economic growth for a Republican governor whose political party controls the government. It also matters for a

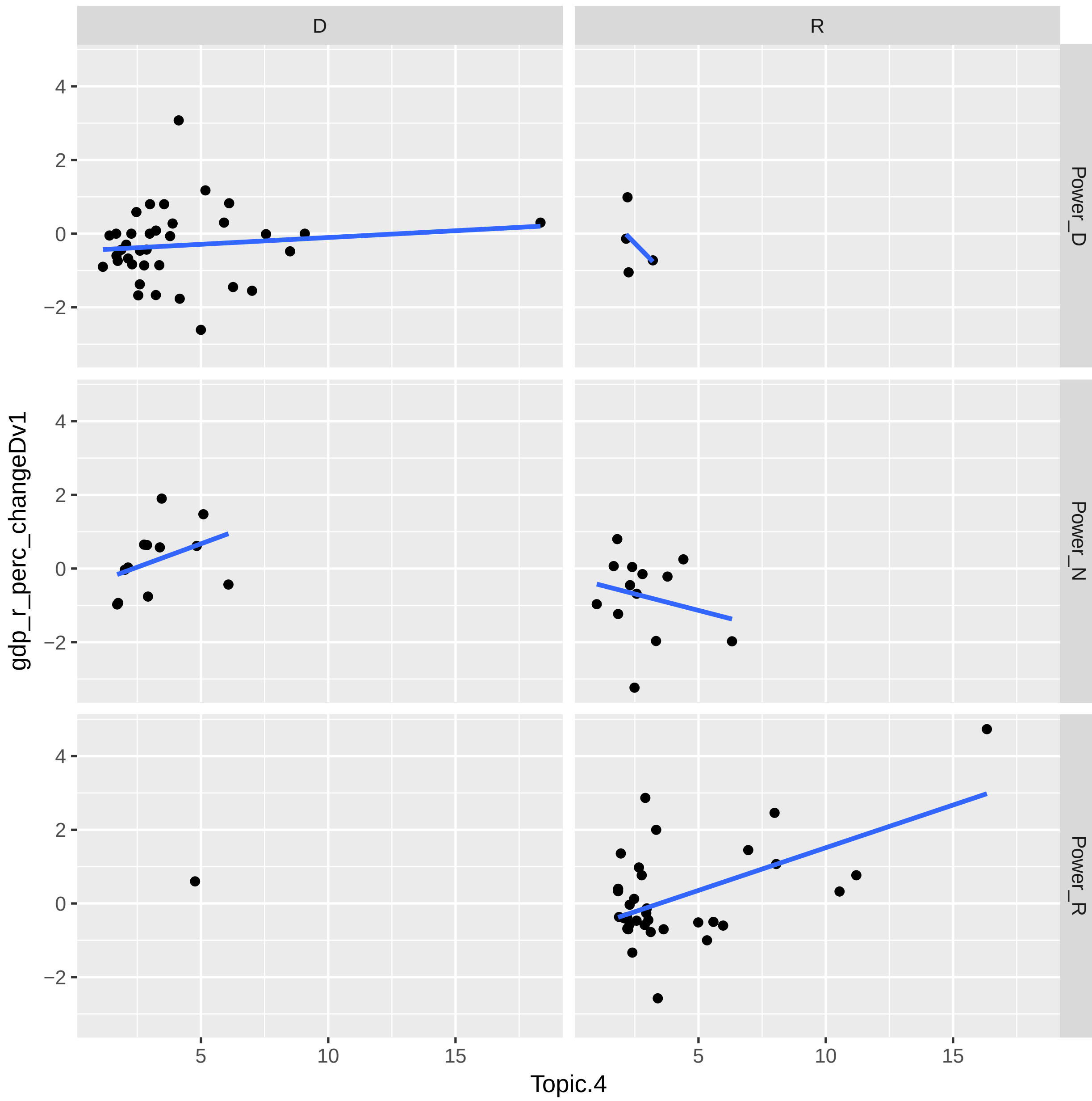


Figure 2: Relationship between the consistency with which governors talk about the economy and the state real GDP average deviation from U.S growth rate. The economic agenda of republican governors whose party control the legislative power tend to achieve higher economic growth than their conterparts. The relationship seems weak for Democrat governors with political power

Table 9: OLS with 5 topics (1) and 2 topics picked by LASSO (2), after removing the two outliers of Topic.4 (3), after taking the log of Topic.4 (4)

	Dependent variable:			
	Two-period lead of state average real GDP growth(deviation from US)			
	(1)	(2)	(3)	(4)
Topic.1	−0.014			
	(0.021)			
Topic.2	0.036			
	(0.040)			
Topic.3	-0.081^{*}			
	(0.044)			
Topic.4	0.110^{***}	0.110^{***}	0.103^{*}	0.477^{**}
	(0.039)	(0.039)	(0.055)	(0.204)
Topic.5	0.055^{*}	0.049^{*}	0.049^{*}	0.429^{*}
	(0.031)	(0.028)	(0.029)	(0.220)
Constant	−0.475	-0.740^{***}	-0.720^{***}	-1.200^{***}
	(0.306)	(0.214)	(0.245)	(0.343)
Observations	102	102	100	102
R^2	0.137	0.102	0.067	0.101
Adjusted R^2	0.092	0.084	0.048	0.083
Residual Std. Error	1.091 (df = 96)	1.096 (df = 99)	1.103 (df = 97)	1.097 (df = 99)
F Statistic	3.054^{**} (df = 5; 96)	5.613^{***} (df = 2; 99)	3.494^{**} (df = 2; 97)	5.561^{***} (df = 2; 99)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 10: Interaction with Governor's Political Affiliation and Ideology. OLS without a special treatment of the outliers (1), OLS after removing the two outliers (2), OLS after taking the log of Topic.4 (3)

	Dependent variable:		
	Two-period lead of state average real GDP growth(deviation from US)		
	(1)	(2)	(3)
Topic.5	0.048*	0.049*	0.379*
	(0.028)	(0.028)	(0.221)
Topic.4:PowerPower_D:PartyD	0.063	0.043	0.334
	(0.046)	(0.065)	(0.223)
Topic.4:PowerPower_N:PartyD	0.241**	0.227**	0.844**
	(0.102)	(0.108)	(0.343)
Topic.4:PowerPower_R:PartyD	0.141	0.130	0.448
	(0.230)	(0.234)	(0.720)
Topic.4:PowerPower_D:PartyR	0.050	0.029	0.268
	(0.230)	(0.237)	(0.669)
Topic.4:PowerPower_N:PartyR	−0.081	−0.096	−0.125
	(0.112)	(0.118)	(0.370)
Topic.4:PowerPower_R:PartyR	0.146***	0.132**	0.593**
	(0.046)	(0.060)	(0.226)
Constant	−0.690***	−0.639**	−1.089***
	(0.227)	(0.256)	(0.351)
Observations	101	99	101
R^2	0.180	0.144	0.172
Adjusted R^2	0.119	0.078	0.109
Residual Std. Error	1.079 (df = 93)	1.090 (df = 91)	1.085 (df = 93)
F Statistic	2.921*** (df = 7; 93)	2.191** (df = 7; 91)	2.751** (df = 7; 93)
Note:			*p<0.1; **p<0.05; ***p<0.01

Democrat governor whose political party do not control the government. A Democrat governor whose political party control the government can affect the economy positively. However, this relationship is not statistically significant. A Republican governor economic agenda effect on economic growth is not statistically significant when his political party does not control the state government.

5.3 Robustness Check

5.3.1 A different consistency measure

A different consistency measure is used for robustness check. That consistency measure is the log of the inverse of the variance. The cross validation method suggests the use of 16 topics when the consistency measure of choice is the log of the inverse of the variance. The Table 11 summarizes the regression results.

Column 1 and 3 shows the results when using all 16 topics. Column 2 and 4 shows the regression results after using a LASSO method to reduce the number of topics. It is worthwhile noting that the topics that are significant when the dependent variable is the one-period lead growth rate are also significant when the dependent variable is the two-period lead growth rate.

Topic.15 is the economic topic when we consider a 16 topics set (Appendix C, Table 13). Topic.15 is significant at 5% significance level for the one-period lead of the growth variable (column 1 and 2), and highly significant for the two-period lead growth variable (column 3 and 4). From Table 13, it can be seen that Topic.15 is very similar to Topic.4 in term of the ranking of their most important words. That confirms that the economic agenda of the governor is positively associated with economic growth. Topic.7, which is about reforms is also positively associated with economic growth. Note that Topic.3 in section 5.1, which has a reform tone to it, was found to be negatively associated with economic growth. There is an apparent contradiction between the finding in section 5.1 and the robustness check finding with respect to the reform topic. But, it is clearer that Topic.7 is about reforms, when Topic.3 in section 5.1 is a mix of several topics with a reform tone. It is also safe to say that Topic.3 in section 5.1 is not specific enough.

Another significant topic is Topic.8, which is clearly about education (see Table 13); it is also positively associated with economic growth. This suggests that the state governor's education agenda is positively associated with economic growth. There are a few canals where state education policy can induce economic growth in the relatively short run. First, it sends a signal to potential investors that the state is serious about improving its education system and company can expect to tap into well-trained workers in the coming years. Second, it helps the state attract talents because highly qualified workers may be concerned about the education their children get. The education topic was found to be positively associated with economic growth in section 5.1 and 5.2, where the

Table 11: OLS with 16 topics (1 and 3); and OLS with topics picked by LASSO (2 and 4)

	Dependent variable:			
	One-period lead (1 and 2) and Two-period lead (3 and 4) of state average real GDP growth(deviation from US)			
	(1)	(2)	(3)	(4)
Topic.1	−0.074		−0.073	
	(0.093)		(0.090)	
Topic.2	−0.072		−0.024	
	(0.072)		(0.069)	
Topic.3	−0.100	−0.119	−0.128	
	(0.091)	(0.085)	(0.087)	
Topic.4	0.063		0.098	
	(0.098)		(0.094)	
Topic.5	−0.041	−0.048	−0.053	−0.083
	(0.093)	(0.079)	(0.089)	(0.076)
Topic.6	−0.061		−0.126	−0.128*
	(0.083)		(0.080)	(0.069)
Topic.7	0.271***	0.244***	0.263***	0.219***
	(0.083)	(0.076)	(0.079)	(0.072)
Topic.8	0.185**	0.182**	0.163*	0.183**
	(0.086)	(0.081)	(0.083)	(0.078)
Topic.9	0.055		0.026	
	(0.079)		(0.076)	
Topic.10	−0.164*	−0.136	−0.071	
	(0.096)	(0.088)	(0.093)	
Topic.11	0.093		0.055	
	(0.109)		(0.104)	
Topic.12	−0.137	−0.118	−0.122	
	(0.084)	(0.077)	(0.080)	
Topic.13	0.048		0.021	
	(0.094)		(0.090)	
Topic.14	0.008		−0.009	
	(0.100)		(0.096)	
Topic.15	0.259**	0.198**	0.329***	0.284***
	(0.107)	(0.096)	(0.102)	(0.093)
Topic.16	0.021		0.045	
	(0.110)		(0.105)	
Constant	−2.490	−1.537	−2.784	−3.514***
	(2.287)	(1.486)	(2.194)	(1.153)
Observations	102	102	102	102
R^2	0.276	0.239	0.310	0.248
Adjusted R^2	0.140	0.182	0.180	0.209
Residual Std. Error	1.082 (df = 85)	1.055 (df = 94)	1.037 (df = 85)	1.019 (df = 96)
F Statistic	2.027** (df = 16; 85)	4.211*** (df = 7; 94)	2.384*** (df = 16; 85)	6.326*** (df = 5; 96)
Note:				*p<0.1; **p<0.05; ***p<0.01

education topic where referred as Topic.2 (see Table 4 column 1, and Table 9 column 1).

In sum, the results obtained using only a five topics construct with the inverse of the coefficient of variation as our consistency measure and the results obtained using a sixteen topics construct with the log of the inverse of the variance are similar. The governor's focus on economic issues matters for economic growth.

5.3.2 Governor's Agenda and state spending

Part of the argument that leadership matters for economic growth is that leaders design, or at least decide on economic, educational, and social policies. The policy options have a bearing on economic outcomes since they affect how society's scarce resources are allotted. For instance, we would expect a governor with an education agenda to devote a high share of state spendings to education. Indeed, education tends to be prominent in U.S governors' state of the state speeches. Of the five topics we constructed, Topic.2, the easiest to interpret of all the five topics, is about education. Does this topic correlate with the share of state spending devoted to education? The following scatter plot shows that the consistency with which governors talk about education is positively correlated with state spending on education. That is, the higher education is a priority for a governor, the higher the share of his/her state spendings devoted to education during that governor's tenure.

The positive relation is valid both for Democrats and Republicans as shown on the graph (Figure 4).[3]

We further plot the correlations of all the topics and the share of state spending on education (Figure 5).

The plots show the correlations and their strength (as measured by their p-values) between the five topics and the education share of the state spendings. The first row of the panel shows the relationship when using annual data (i.e the annual speech is the unit of observation). The second row of the panel shows the relationship when using the constructed consistency measure (i.e, the governor is the unit of observation). The bars in the first column represent the correlation values, and the bars in the second column represent the p-values of the correlations. Topic.2 (the education topic) is emphasized with a darker bar. The first row of the plot suggests a strong relationship between what the governor says in his/her speech (at the beginning of the year) and the share of the state spendings devoted to education (in that year).

The second row suggests that the governor consistency over the education topic is fairly correlated, positively, with the average share of state spending on education.

[3] We dropped the observations for which the education consistency measure is greater than 6 to highlight the spread of the observations. By dropping the 9 observations with values greater than 6, the qualitative trend of the data does not change.

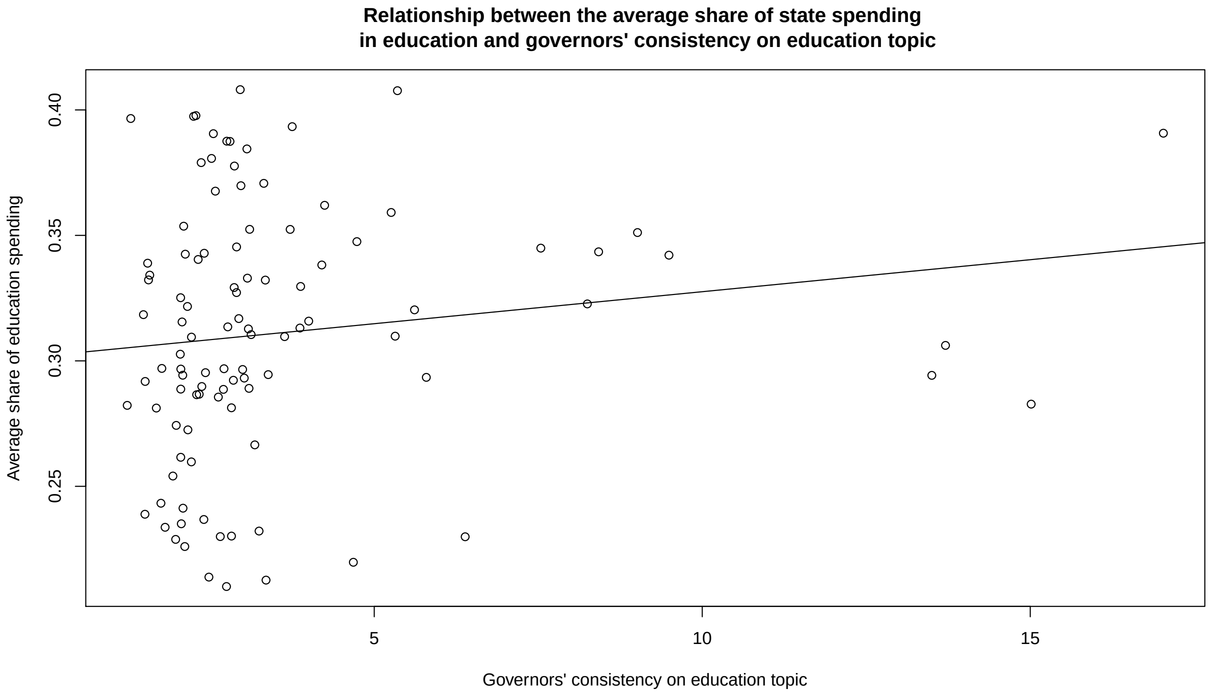


Figure 3: The consistency with which governors talk about education is positively correlated with the average share of state spendings on education.

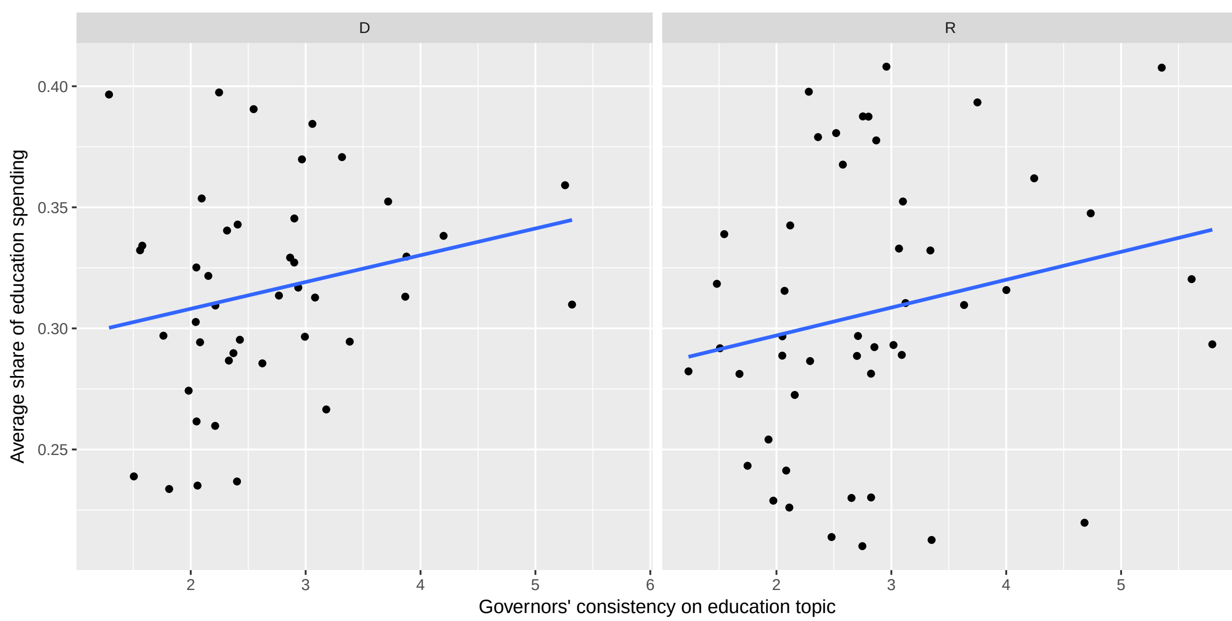


Figure 4: The consistency with which governors talk about education is positively correlated with the share of state average spendings on education. The relationship is similar both for Democrats and Republican governors.

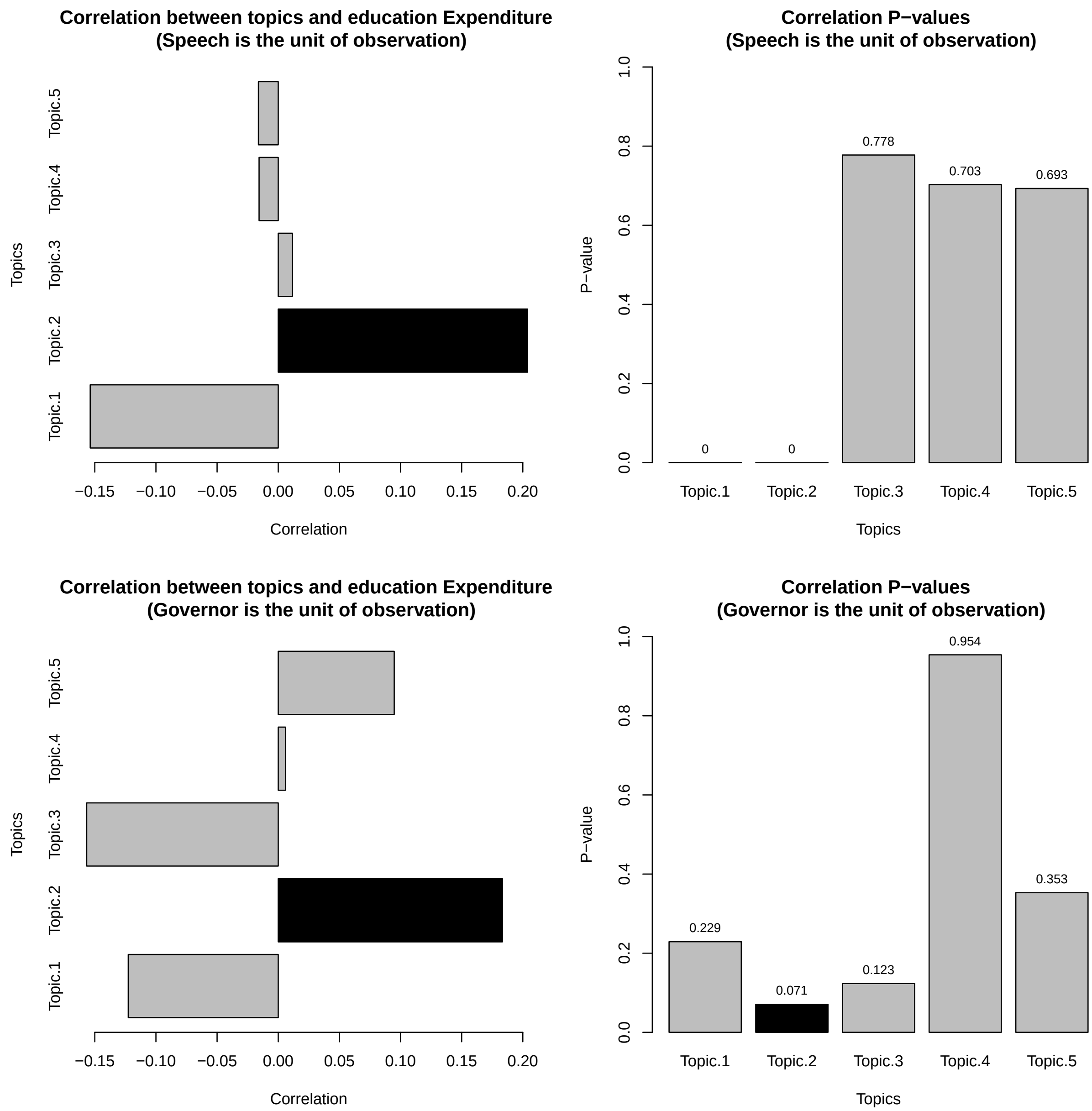


Figure 5: The correlation between the importance of the topics and the share of state spending on education (first row) is highest for the education topic (Topic.2). The correlation is highly significant with a p-value close to zero. Using the constructed variables, where the governor is the unit of observation, the education topic has the highest correlation value (second row). The correlation is marginally significant, with a p-value = 0.071.

For both cases (i.e. using annual speeches as the unit of observation, or state governor as the unit of observation), the education topic is the most correlated with the share of state spendings on education. That suggests that the governor's education agenda matters for the resources the state devotes towards education[4]

.

5.4 Goal driven leaders and economic growth

The main take away from the robustness check is that the relevant topics found using the inverse of the coefficient of variation as the consistency measure are also relevant when we use the log of the inverse of the variance as the consistency measure. This suggests that being consistent on certain topics matters for economic growth; however we measure consistency. This finding supports the main thesis of this paper, which postulates that the goal-driven leader is persistent on his/her agenda; and persistence yields results. To capture the persistence of the governor, we constructed a consistency measure which captures governors' consistency over certain topics. Being persistent on a topic over several years is a sign of commitment to an agenda. An example of a commitment to an agenda can be seen by looking at the speeches of the governor of West Virginia from 2001 to 2004[5]. It is safe to say that the governor is an education governor. We arrive at this conclusion because this governor scores high on our education consistency measure. Then, by looking at his speeches over four years period, it is apparent that this governor main focus was in improving West Virginian education system. Another example is the North Dakota governor of 2001 to 2009 whose economic development agenda rank among the top in our consistency measure[6].

5.5 Text analytics and leadership studies

The main goal of the present paper was to study US governors' professed agenda and economic growth. We define professed agenda the consistency with which governors talk about certain topics in their state of the state speeches. Talking about a topic consistently is perceived as a sign of commitment to an agenda. Thus, we first define a consistency measure which is the inverse of the coefficient of variation ($C_{i,j} = \frac{\bar{X}_{i,j}}{s_{i,j}}$). The main idea of this definition is that high variance is a sign of lack of focus, and high mean is a sign of importance. The means and variances are computed from the proportion of the speeches devoted to a particular topic. By defining consistency as the inverse of the coefficient of variation, a high score indicates that the governor shows low variability and high proportion of a given topic in his/her speeches.

[4]It should be noted that four outliers valued where dropped before computing and plotting the correlations seen on the second row of the panel.

[5]http://stateofthestate.com/content.aspx?state=WV&date=02/14/2001

[6]http://stateofthestate.com/content.aspx?state=ND&date=01/09/2001

One main challenge of text analytics is separating signals from noise. We rely on cross-validation methods to identify a limited set of topics that are correlated with economic growth rate. It turns out that the identified topics bear economic significance. For example, we found that the economic development topic is always statistically, significantly related to economic growth for all the different specifications we used. Using a different definition of the consistency measure, the log of the inverse of the variance, the topics that are statistically related to economic growth are those expected when indeed the governors do what they profess to be doing. Furthermore, the same topics were identified when using the inverse of the coefficient of variation as the consistency measure. These findings suggest that there are signals in the governors' speeches. The findings further show that text analytics methods applied to leaders' speeches can provide measures of leaders' professed agenda. This provides an avenue for studying political leadership and economic growth. We do not have a measure of what they do, but we can measure what to say they are doing. Consequently, this paper makes the case that text analytics, applied to leaders' speeches, is a viable approach to studying political leadership and economic growth.

6 Conclusion

This paper demonstrates that text analytics methods, particularly topic modeling, applied to leaders' speeches, is a viable tool for studying political leadership and economic performance. Using a corpus of 598 U.S governors' state of the state speeches, topic modeling allowed us to identify 5 topics covered in these speeches. Of the 5 topics, the LASSO method was used to identify the topics which co-vary most with real GDP growth rate (state growth average deviation from U.S growth rate). Then, simple OLS regression using the most important topics suggested that only one topic (the economic development topic) matters for economic growth under the diverse specifications used in this paper.

The paper demonstrates the possibility of studying U.S governors in terms of what they say, rather than their ideological affiliation to a political party, as is often seen in the economics literature (Beland (2015)).

The paper used mainly 5 topics for a corpus of 598 speeches. Five topics, though enough for the current paper's purpose, may appear too small. For example, Topic.3 appears to be a mix of several topics; and had we chosen a larger set of topics (maybe 10) this mix-up topic could be separated to well-defined topics. Indeed, by redefining the consistency measure as the log of the inverse of the variance, the reform topic (which is apparent in Topic.3 when using 5 topics set) appeared as a meaningful topic. The governor's education agenda is also positively related to economic growth. From our findings, it appears that the governor's economic development, education, and reform agendas are strongly associated with economic growth. A future research would be to look at governor's professed agendas and action taken by the state to achieve these agenda. For instance, are governor's with professed education agenda spend more on education? Are governor's with reform agenda pass more reforms legislation?

References

Alexopoulos, M. and Cohen, J. (2015). The power of print: Uncertainty shocks, markets, and the economy. *International Review of Economics & Finance*, (0):–.

Baker, S. R., Bloom, N., and C, S. J. D. (2013). Measuring economic policy uncertainty.

Beland, L.-P. (2015). Political parties and labor-market outcomes: Evidence from us states. *American Economic Journal: Applied Economics*, 7(4):198–220.

Berry, W. D., Ringquist, E. J., Fording, R. C., and Hanson, R. L. (1998). Measuring citizen and government ideology in the american states, 1960-93. *American Journal of Political Science*, 42(1):327–348.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

Brady, D. and Spence, M. (2010). *Leadership and Growth : Commission on Growth and Development.* World Bank., https://openknowledge.worldbank.org/handle/10986/2404.

Byman, D. L. and Pollack, K. M. (2001). Let us now praise great men: Bringing the statesman back in. *International Security*, 25(4):pp. 107–146.

Einav, L. and Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210).

Grimmer, J. (2010). A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, 18(1):1–35.

Hermann, M. G. (2008). *Using Content Analysis to Study Public Figures.* Palgrave Macmillan.

Hermann, M. G., Preston, T., Korany, B., and Shaw, T. M. (2001). Who leads matters: The effects of powerful individuals. *International Studies Review*, 3(2):83–131.

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, UAI'99, pages 289–296, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Mach. Learn.*, 42(1-2):177–196.

Jing, F. F. and Gayle (2008). Missing Links In UnderstandingThe Relationship Between LeadershipAnd Organizational Performance. *International Business & Economics Research Journal*, Volume 7(Number 5).

Jones, B. F. and Olken, B. A. (2005). Do leaders matter? national leadership and growth since world war II. *The Quarterly Journal of Economics*, 120(3):835–864.

Landauer, T. K., Mcnamara, D. S., Dennis, S., and Kintsch, W., editors (2007). *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates.

LAVER, M., BENOIT, K., and GARRY, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, null:311–331.

Lieberson, S. and O'Connor, J. F. (1972). Leadership and organizational performance: A study of large corporations. *American Sociological Review*, 37(2):pp. 117–130.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., and Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1):209–228.

Romer, C. D. and Romer, D. H. (2015). New Evidence on the Impact of Financial Crises in Advanced Countries. NBER Working Papers 21021, National Bureau of Economic Research, Inc.

Shlens, J. (2014). A tutorial on principal component analysis. *CoRR*, abs/1404.1100.

Solka, J. L. (2008). Text data mining: Theory and methods. *Statist. Surv.*, 2:94–112.

Thomas, A. B. (1988). Does leadership make a difference to organizational performance? *Administrative Science Quarterly*, 33(3):pp. 388–400.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28.

Wilkerson, J., Smith, D., and Stramp, N. (2015). Tracing the flow of policy ideas in legislatures: A text reuse approach. *American Journal of Political Science*, pages n/a–n/a.

Winter, D. G. (2005). Things i've learned about personality from studying political leaders at a distance1. *Journal of Personality*, 73(3):557–584.

Zubin, J., Bruce, K., and Suresh, N. (2014). Detecting latent ideology in expert text: Evidence from academic papers in economics. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1804–1809. Association for Computational Linguistics.

A Topic Modeling

A.1 Topic models

Topic modeling is a refinement of Probabilistic Latent Semantic Allocation (PLSA), which is itself a refinement of the Latent Semantic Allocation (LSA) model, which is easily understood through Principal Component Analysis (PCA). The unifying theme of all of these methods is to reduce the dimension of the data, and hopefully, we can say something interesting about the original data, by looking at the reduced form of the data. Clearly, a scatter-plot can tell us a lot when we are dealing with two variables. Beyond two variables, scatter-plot becomes insufficient. Maybe there is a way to collapse several variables into, say two variables, and still preserve the most important features of the original data. This is what these methods try to accomplish. Analogically, these methods can be perceived as changing your standing position in the hope of seeing things one couldn't see at his/her initial position. To understand what topic modeling achieves, it is helpful to understand PCA.

A.1.1 Principal Component Analysis

Principal component analysis is one of the most popular statistical methods, which aims at reducing data dimensions. Assume we have a data set with tens of variables, some of which are correlated, and a few of them are driving the variations across observations. The goal of PCA is to re-express the data in such a way redundancies between variables are eliminated, and variables' variances are highlighted. This goal is achieved by projecting (rotating to be more exact) the data into a new basis in which the variables are no longer correlated.

Let assume $X_{n,p}$ is our matrix of variables. Let's further assume $x_{n,p}$ to be the standardized format of the X variables, that is $x_{n,p} = \frac{(X_{n,p}-\bar{X})}{\sigma_X}$[7]. It is easily seen that the dot product, $x^t_{n,p}x_{n,p} = A_{p,p}$, is the correlation matrix of the initial data, where $x^t_{n,p}$ is the transpose of $x_{n,p}$. By eigenvalue decomposition, $A_{p,p}$ can be expressed in this format:

$$A_{p,p} = U_{p,k}D_{k,k}U^t_{k,p}$$

[7] The $x_{n,p}$ are the z-values of each variable p

where U is an orthogonal matrix (i.e $U^tU = I$), and D is a diagonal matrix, with the eigenvalues its diagonal elements. Note that this expression is just a mathematical identity. It turns out that $U^t$is a rotation matrix (a new basis to re-express the data); $y_{n,p} = U^t_{n,p}x_{n,p}$ is the re-expressed data in the new basis. $y_{n,p}$ is a matrix of uncorrelated variables. Further, it turns out that D is the variance matrix of $y_{n,p}$. To see why that is the case, let's assume that there is an orthogonal matrix U such that:

$$y = xU,$$

so,

$$y^ty = U^tx^txU = U^tAU$$

Knowing that $A = UDU^t$, we can see that:

$$y^ty = U^tx^txU = U^tAU = U^tUDU^tU = D,$$

so y^ty is the diagonalized form of x^tx. In other words, the re-expressed x is y, and $y^ty = D$ shows that the covariance between the y variables are zeros. The diagonal elements are written in descending order, and since the diagonal elements are the variances, the hope is that the first few k diagonal elements (variances) represent a "large" percent of the total sum of the variances, suggesting that the remaining $p-k$ y variables can be dropped without loss of relevant information in the original data. Hence, the idea of dimensionality reduction. Also, note that if the original data is not correlated, then the correlation matrix is already diagonal, and PCA becomes useless. Shlens (2014) has an impressively accessible tutorial on PCA.

A.1.2 Latent Semantic Analysis

LSA was the state of art of text analystics in the 90s. It is an application of the Singular Value Decomposition (SVD) to the matrix of text data (called document-term matrix (DTM)), the goal being to reduce the dimensionality of the DTM, which is a sparse matrix. Table 1 shows an example of a DTM. Notice the presence of too many cells with zero counts. SVD allows the conversion of the DTM into a smaller dimension, in PCA like fashion. In fact, It can be shown that SVD is somewhat identical to PCA.
Let M be the DTM. M can be written as:

$$M = U\Sigma V^T,$$

where U and V are orthogonal matrices (i.e. $UU^T = VV^T = I$), and Σ is a diagonal matrix.
LSA consists of finding the matrices U, V, and Σ. To understand what LSA does, it suffices to see

its relation to PCA. From the above SVD decomposition, we have:

$$M^t M = [U\Sigma V^T]^t [U\Sigma V^t] = V\Sigma\Sigma V^t$$

Note that this equation is identical to the PCA equation, when we set $D = \Sigma\Sigma$. Similarly, we can write

$$MM^t = [U\Sigma V^T][U\Sigma V^T]^t = U\Sigma\Sigma U^t$$

Hence, LSA is very similar to PCA, except that LSA provides results that assume observations to be variables too. It is important to see how LSA is very similar to PCA, because topic modeling is just a variant of LSA; and topic modeling is so mathematically involved that it is easy to lose sight of what it does.

$M^t M$is a similarity measure. M being the matrix of words count with its rows the documents and its columns the words, $M^t M$ is similar to a covariance matrix. Thus, by rotating the matrix M with the rotation matrix U, we hope to capture the relationship between document in a new basis, in which the most important features in these documents are preserved.

Similarly, MM^t is a similarity matrix (words similarity). By assuming documents to be the variables, and by rotating M^t with the matrix$V$ we hope to capture the similarity between words. Words that are very close to each other are hopefully related to a distinguishable topic. A further exposition of LSA can be found in Landauer et al. (2007), chap1&2.

A.1.3 Probabilistic Latent Semantic Allocation

Although, the LSA has been largely used, in several domains (lingustics, psychology, computer science,...) it has some shortcomings. First, the rotated data cannot be interpreted since its elements take any kind of numbers (negative and positive values); second, the choice of K is arbitrary; and third, it does not assume any data generative process, which prevent inferential application of the results (Hofmann, 1999, 2001). Probabilistic Latent Semantic Allocation (PLSA), introduced by Hofmann (1999), addresses these problems, by bringing probability distributions into LSA models. PLSA can be loosely expressed as follows:

$$M = \theta_{D,K}\phi_{K,V},$$

where M is a matrix of words distribution over documents, θ is a matrix of topics distribution over documents, and ϕ is a matrix of words distribution over topics.

To see how PLSA is conceptually related to LSA, it suffices to note that the Σ matrix of LSA can be absorbed in the matrix U, or V. Thus, we can write:

$$M = U\Sigma V^T = U\tilde{V},$$

where $\Sigma V^t = \tilde{V}$. Maximum Likelihood methods are used to estimate the matrices $\theta_{D,K}$, and $\phi_{K,V}$, simultaneously.

It is worthwhile repeating that this paper uses the topic modeling method; however, it is easier to grasp the method if we understand its link with PCA, which is intuitively easier to picture in the mind.

PLSA can be seen as classical statistics application to the LSA model. LSA is purely mathematical (matrix algebra). PLSA brings probability distribution assumption into LSA. Topic modeling is just a Bayesian extension of PLSA.

A.1.4 Topic modeling

In PLSA, θ and ϕ are assumed to be parameters of multinomial distributions, then by using a maximum likelihood estimation method, we can estimate θ, ϕ, and K (K is selected based on model comparisons). Assuming that the parameters θ and ϕ are themselves random variables with Dirichlet prior distribution, Blei et al. (2003) proposes the Latent Dirichlet Allocation (LDA), which is a bayesian approach to PLSA. LDA is generically refers as topic modelling. Following is an exposition of the model.

$$\begin{array}{c} \\ d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_D \end{array}
\begin{array}{c} \begin{array}{cccc} w_1 & w_2 & \cdots & w_V \end{array} \\ \begin{pmatrix} n_{1,1} & n_{1,2} & \cdots & n_{1,V} \\ n_{2,1} & n_{2,2} & \cdots & n_{2,V} \\ n_{3,1} & n_{3,2} & \cdots & n_{3,V} \\ \vdots & \vdots & \vdots & \vdots \\ n_{D,1} & n_{D,2} & \cdots & n_{D,V} \end{pmatrix} \end{array}
\approx
\begin{array}{c} \\ d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_D \end{array}
\begin{array}{c} \begin{array}{cccc} t_1 & t_2 & \cdots & t_K \end{array} \\ \begin{pmatrix} \theta_{1,1} & \theta_{1,2} & \cdots & \theta_{1,K} \\ \theta_{2,1} & \theta_{2,2} & \cdots & \theta_{2,K} \\ \theta_{3,1} & \theta_{3,2} & \cdots & \theta_{3,K} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{D,1} & \theta_{D,2} & \cdots & \theta_{D,K} \end{pmatrix} \end{array}
*
\begin{array}{c} \\ t_1 \\ t_2 \\ \vdots \\ t_K \end{array}
\begin{array}{c} \begin{array}{cccc} w_1 & w_2 & \cdots & w_V \end{array} \\ \begin{pmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,V} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,V} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{K,1} & \phi_{K,2} & \cdots & \phi_{K,V} \end{pmatrix} \end{array}$$

The first matrix above is a matrix of count data. That is, for a given document d, $n_{d,v}$ is the number of time the word w_v is used in the document d. This matrix is called Document-term matrix (DTM). The second matrix shows the distribution of topics per document; that is, a document is a collection of topics with varying proportions. These distributions are $\theta_{d1}, \theta_{d2}, \theta_{d3}, ..., \theta_{dK}$, where $\sum_{k=1}^{K} \theta_{dk} = 1$, so for a given document, we have:

$$z_d|\theta \sim multinomial(\theta),$$

that is,

$$p(z_d|\theta) = \prod_{k=1}^{K} \theta_{d,k}^{n_{d,k}}$$

where z is a vector of topics, and $n_{d,k}$ is the number of words that characterize topic k in document d. Further, if we assume that documents are independent, then the joint topics distribution of the corpus (the collection of documents) is:

$$p(z|\theta) = \prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{d,k}^{n_{d,k}}$$

The third matrix shows the distribution of words per topic. A topic is a distribution over words, that is, a topic is a collection of words with variant frequencies. These distributions are ϕ_{k1},ϕ_{k2} , ϕ_{k3},...ϕ_{kV}, where $\sum_{v=1}^{V} \phi_{k,v} = 1$, so for a given document, we have:

$$w|\phi \sim multinomial(\phi),$$

that is,

$$p(w_k|\phi) = \prod_{v=1}^{V} \phi_{k,v}^{n_{k,v}},$$

and assuming topic independence, the joint topic distribution over words is given by:

$$p(w|\phi, z) = \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{k,v}^{n_{k,v}},$$

where $n_{k,v}$ is the count of the number of times topic k was assigned to the word v in the whole corpus.

The above probability assumptions gave rise to the PLSA Hofmann (1999). By further assuming that the parameters θ and ϕ are random variables following, respectively, $Dirichlet(\alpha)$, $Dirichlet(\beta)$, Blei et al. (2003) created the Latent Dirichlet Allocation, the so called LDA. So, we have:
for a given document,

$$p(\theta_d|\alpha) = \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \theta_{d,1}^{\alpha_1 - 1} ... \theta_{d,k}^{\alpha_k - 1} ... \theta_{d,K}^{\alpha_K - 1},$$

and for the whole corpus,

$$p(\theta|\alpha) = \prod_{d=1}^{D} \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \theta_{d,1}^{\alpha_1 - 1} ... \theta_{d,k}^{\alpha_k - 1} ... \theta_{d,K}^{\alpha_K - 1},$$

and for a given topic,

$$p(\phi_k|\beta) = \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \phi_{k,1}^{\beta_1 - 1} ... \phi_{k,v}^{\beta_v - 1} ... \phi_{k,V}^{\beta_V - 1},$$

and for the whole set of K topics,

$$p(\phi|\beta) = \prod_{k=1}^{K} \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \phi_{k,1}^{\beta_1 - 1} ... \phi_{k,v}^{\beta_v - 1} ... \phi_{k,V}^{\beta_V - 1}$$

Collecting all the probabilities above, the joint distribution is:

$$p(w,z,\theta,\phi|\alpha,\beta) = p(\theta|\alpha)p(z|\theta)p(\phi|\beta)p(w|z,\phi)$$

z, θ, and ϕ, are latent variables. By marginalizing $p(w,z,\theta,\phi|\alpha,\beta)$ over the latent variables, we have:

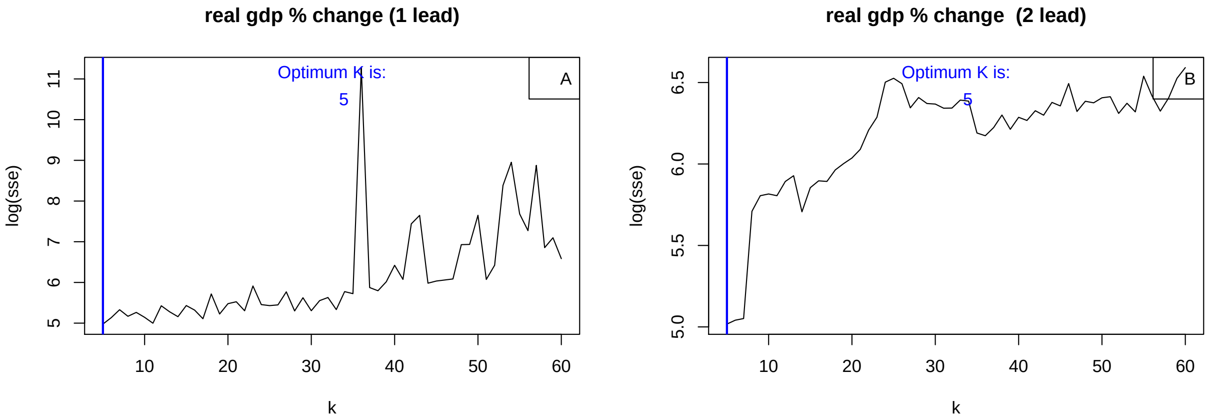
$$p(w|\alpha,\beta) = \int_\theta \int_\phi \sum_z p(\theta|\alpha)p(z|\theta)p(\phi|\beta)p(w|z,\phi)$$

The overarching goal of the LDA is to estimate θ and ϕ. Again, the matrix of $\theta_{d,k}$ are the topics k's distribution in document d, and the matrix of $\phi_{k,v}$ are the words distributions for topic k. Thus the value $\theta_{d,k}$ informs on the importance of topic k in document d, and $\phi_{k,v}$ informs on the importance of word v in topic k. In practice, the first thirty most important words are often used to infer the meaning of the topic.

The usefulness of topic modeling for the current paper is its ability to automatically provide the topics distributions $\theta_{d,k}$, d being a speech. Put differently, topic modelling informs on the topics, and their relative importance in every leader's speech. Knowing the relative importance of each topic in each leader's speech, and how the importance of topics changes over time can inform on the priorities of a leader.

B Choosing the optimum K

Topic modeling algorithms require that the number of topics K be provided. To avoid picking K arbitrarily, we use a ten folds cross-validation approach to picking the optimum K; that is, we iterate the regression through different values for K. For each K, an OLS is run on 90% of the observations; then the 10% remaining observations are used to predict the remaining values for the dependent variable. Using the predicted values, we compute the errors sum of squares (sse) of the dependent variable for the 10% of the observations. The process is repeated 10 times, each using a different set of observations to be predicted. For each K, a single value for the sse is computed. Then the K for which the sse is small is considered the optimum number of topics to be used.



The figure above shows the path of the log sses as K takes different values ($K = 5, 6, ..., 60$). The vertical blue line indicates the optimum number of topics K. K = 5 is the optimum K for the two dependent variables. The dependent variables are the state average deviation from the U.S growth rate of the real GDP (one-period lead, and two-period lead). Each independent variable is a consistency measure over a topic. We measure consistency as the inverse of the coefficient of variation (CV). We are interested in governors who talk a lot about a topic (which we capture by the mean $\bar{X}_i$), with a minimum variance over time (which we capture by the standard deviation s_i). Thus the inverse of the CV, that is $\bar{X}_i/S_i$ can be seen as a consistency measure.
Note: the observational unit is a governor.

C Topics' words table

The topics' words table are often used to interpret the topics. The words are ordered in terms of their relative importance (proportion) with respect to the topic. The following table shows the keywords by topic. A more elaborate tool for interpreting the topics can be found at this link: `https://salif.shinyapps.io/topic_context` Topic.1 and Topic.3 remains diffused and difficult to interpret.

D Web links of economic speeches

`http://www.americanrhetoric.com/speeches/gwbusheconomicgrowthpackage.htm`
`http://www.presidentialrhetoric.com/speeches/02.02.06.html`
`http://georgewbush-whitehouse.archives.gov/news/releases/2003/01/20030107-5.html`
`https://www.whitehouse.gov/the-press-office/2012/06/14/remarks-president-economy-clev`

Table 12: Topics' words table

	Topic.1	Topic.2	Topic.3	Topic.4	Topic.5
1	state	school	will	will	state
2	peopl	educ	state	state	fund
3	year	year	must	work	will
4	want	student	govern	busi	year
5	know	children	time	creat	budget
6	just	teacher	make	energi	million
7	work	health	budget	develop	increas
8	thing	care	futur	also	propos
9	make	everi	work	nation	also
10	need	help	challeng	build	program
11	come	make	econom	compani	servic
12	time	will	economi	help	revenu
13	like	program	face	invest	provid
14	last	famili	reform	economi	public
15	south	high	public	peopl	spend
16	money	invest	today	opportun	continu
17	great	colleg	citizen	econom	need
18	thank	state	famili	make	percent
19	good	work	busi	industri	addit
20	back	need	togeth	year	dollar
21	think	last	protect	futur	educ
22	said	parent	need	provid	current
23	look	learn	nation	thank	feder
24	take	child	everi	communiti	legisl
25	chang	system	governor	effort	support
26	govern	communiti	peopl	west	depart
27	mani	first	educ	technolog	govern
28	first	provid	respons	plan	includ
29	give	insur	servic	million	project
30	right	nation	live	need	plan
31	believ	improv	leader	grow	issu
32	bill	better	serv	must	develop
33	import	must	come	home	system
34	better	qualiti	mani	univers	last
35	that	propos	children	like	cost
36	talk	read	great	north	administr
37	start	rais	system	resourc	address
38	much	opportun	address	today	improv
39	everi	life	choic	center	local
40	spend	increas	health	product	recommend

Table 13: Topics' words table

	Topic.3	Topic.5	Topic.6	Topic.7	Topic.8	Topic.10	Topic.12	Topic.15
1	peopl	must	energi	reform	school	year	health	busi
2	know	educ	will	state	educ	state	care	will
3	thing	today	also	govern	teacher	last	children	state
4	just	futur	north	system	student	nation	insur	creat
5	want	opportun	develop	spend	children	program	famili	compani
6	come	invest	creat	school	everi	percent	cost	economi
7	make	innov	peopl	will	need	million	must	year
8	time	work	million	chang	will	public	year	worker
9	think	continu	build	need	learn	increas	work	invest
10	work	univers	renew	cost	read	sinc	make	peopl
11	said	commit	power	year	high	rate	afford	workforc
12	great	growth	percent	taxpay	help	first	help	work
13	like	togeth	compani	incom	classroom	three	protect	help
14	need	build	plant	properti	parent	made	provid	econom
15	back	qualiti	product	billion	child	even	will	make
16	give	effort	help	propos	teach	next	access	train
17	right	communiti	futur	save	grade	time	need	colleg
18	take	challeng	nation	educ	graduat	four	reduc	employ
19	tell	economi	produc	better	better	today	medic	small
20	money	success	wind	competit	take	still	system	high
21	look	life	opportun	busi	first	report	citizen	industri
22	that	research	industri	administr	program	higher	live	skill
23	start	prosper	just	pass	provid	histori	drug	grow
24	talk	center	generat	choic	improv	thank	expand	keep
25	today	provid	plan	first	math	past	nation	build
26	someth	make	electr	lower	colleg	everi	program	everi
27	chang	colleg	school	growth	mani	month	legislatur	next
28	them	ensur	need	local	system	number	child	manufactur
29	world	state	increas	good	propos	progress	senior	just
30	mean	live	fund	long	standard	anoth	communiti	across
31	littl	great	fuel	make	make	done	coverag	bring
32	good	grow	state	pension	time	well	healthi	need
33	here	higher	home	privat	money	said	mani	like
34	place	technolog	market	rais	work	rank	made	move
35	this	econom	provid	real	good	mani	prescript	give
36	sure	also	resourc	control	account	high	home	rate
37	import	best	research	plan	test	special	allow	good
38	believ	strong	take	rate	earli	enact	give	focus
39	live	vision	center	district	rais	pass	increas	compet
40	life	world	project	mani	meet	initi	challeng	technolog

https://www.whitehouse.gov/the-press-office/2013/12/04/remarks-president-economic-mob

http://www.americanrhetoric.com/speeches/barackobama/barackobamacarnegiemellon.
htm

http://www.p2016.org/bush/bush090915sp.html

http://mittromneycentral.com/speeches/2012-speeches/102612-remarks-on-the-american-ec

http://www.marketwatch.com/story/full-text-of-sen-mccains-economy-speech