

# The Governor as the Entrepreneur in Chief: An Exploratory Analysis

Salfo Bikienga

June/2017

## Abstract

Attracting and expanding businesses in the state often appear prominently in the speeches of political leaders. Do they talk the talk or do they actually follow through? Do their commitments to promote businesses matter? The talks about promoting businesses suggest that political leaders play a role in the expansion of business establishments. Though that may be true, the economics literature is silent on the ability of political leaders to alter the dynamics of business establishments. Applying machine learning algorithms to U.S governors' State of the State Addresses from 2001 to 2013, this paper captures the level of the governors' professed business friendly agendas; then studies the relationship between the governor's long term commitment to his/her business agenda and business dynamics in his/her state. The paper shows that the commitment of the governor to expand business establishments in his/her state is positively associated with the growth rate of business establishments in their states. (**JEL O10, R50, C38**)

## 1 Introduction

The literature on development economics and regional economics has emphasized the role of certain economic policies in promoting economic growth. Among others, there is a consensus that human capital formation (Gennaioli et al., 2013), physical infrastructure (e.g., roads, power and internet access) (Romp and De Haan, 2007), access to credits (Kerr and Nanda, 2009) facilitate the emergence of a dynamic and prosperous private sector. Moreover, it has been argued that entrepreneurial projects are causal determinant of future growth (Bunten et al., 2015), with the implication that states should be actively pursuing business expansion. The economics literature is replete of policy recommendations, many of which have been found to be working, theoretically and empirically. The question is why the policies that work are not implemented by all states, and why the same policies work differently in different places. One possible argument is that policies are adopted by policy makers, and the governor is the policy maker in chief at the state level. The governor believes and/or policy preferences may dictate the kind of economic policies that are implemented, and to what extent the state is committed to implementing these policies.

The current paper proposes using governors' State of the State Addresses (SoSA) to capture what they profess to be their agendas and assess the relationship between these professed agendas and the entrepreneurial dynamics in their states. The aim of the paper is to explore whether the

governors' professed agendas matter for entrepreneurial dynamics. To do so, we use the Latent Dirichlet Allocation (LDA) algorithm to quantify the main themes covered by the US governors in their annual SoSA. The quantified themes are then used to construct a consistency measure, which is assumed to capture the level of commitment of the governor to his/her professed agenda.

The idea of using the consistency with which a governor talk about an issue as a measure of policy commitment derives from Hermann (2008); Hermann et al. (2001). Indeed, Hermann (2008) argues that political leaders public statements can be used to study their leadership styles (for example, the goal driven leaders are persistent in what they say, whereas the opportunistic leaders tend to respond to news, and find it difficult to have a consistent message over time). We hypothesize that to get something significant done, a certain level of commitment of the state is necessary. A governor who is committed in promoting business expansion in the state should show signs of commitment through his/her most important speech of the year; the SoSA.

The use of political speech to study political leaders is widely used and accepted in other social sciences. Winter (2005) notes that the "one kind of data from political leaders that is produced and preserved in abundance" is their words. Political leaders communicate their agenda, mobilize followers, and research suggests that their public statements reflect what they wants, and what they are pledging to be (Hermann, 2008). Grimmer (2010) identifies the expressed agendas of the U.S senators using the senate Press Releases.

In practice, we identify the thematic contents of governors' speeches using text analytics and machine learning techniques. These techniques are gaining acceptance in the mainstream economics literature, (Romer and Romer (2015), Varian (2014), Einav and Levin (2014), Mullainathan and Spiess (2017), Athey and Imbens (2017), Baker et al. (2013), Alexopoulos and Cohen (2015)).

The use of text analytics offers an opportunity to study economic questions previously thought too nebulous to approach rigorously. Political leadership is an example of such an interesting, but difficult question to study with traditional structured data (Brady and Spence, 2010).

The overarching goal of the current paper is to explore the relationship between the US governors commitment to attracting businesses in their states and the dynamics of business establishment. The paper shows that the U.S. governors' professed commitment to attracting businesses translate into action that increases the number of businesses in their states. One advantage of topic modeling is our ability to identify the specific policies undertaken by the governors in their quest to attract businesses, since these policy option are stated in their speeches.

The remaining of the paper presents the data and method of analysis in Section 2, followed by an exposition of the results in Section 3. We discuss the main findings and their implication in Section 4, and conclude in Section 5.

## 2 Methodology and Data

This paper uses topic modeling algorithm, the Latent Dirichlet Allocation (LDA) method (Blei et al. (2003), Blei (2012)) to identify the major themes covered in the US governors' SoSA; then uses a linear regression method to study the relationship between the governors' commitment to their agendas (as measured by the persistence with which they talk about issues) and the growth rate of business establishments in their states. The goal is to explore whether the commitment of a governors to the economy promotes the establishment of news businesses in their states.

## 2.1 The Model

Our end goal in this paper is to use a simple linear regression model to study the relationship between the governors commitment to their “professed agenda” and the rate of entry of new establishments in their states. To do so, we construct a consistency variables, which are then used as exogenous variables explaining the establishments entry rate at state level. It can be argued that the contents of the speech is a proxy for the actions the governors intend to undertake [?]. And since it takes time for the actions taken to influence the economy, we use future business establishment entry rates (one year, two years, and three years latter) as outcome variables.

For now, let’s assume we can measure the relative importance of each topic in each of the governor speeches. For instance, we can arrive at numbers such as 60% percent of the speech is about education, 15% is about the economy, 9% is about healthcare and 16% is about others. Then for each governor with several annual speeches, we can construct some measure of consistency for each of the topic covered in his/her speeches.

Let’s  $C_{i,j}$  be the consistency measure of governor  $i$  over the topic  $j$  (how consistency is measured is explained in Section 2.2);

$E_i$  is the average establishment rate of governor  $i$ ’s state during governor  $i$ ’s term. The unit of observation is a governor tenure term, a term being a four year. Therefore, a governor in office for eight years is assumed to be two governors. We chose a term as the unit of observation because political leaders tends to shift their focus from term to term.

The basic model regresses the entry rate on the consistency over topics as follows:

$$E_i = \beta_0 + \sum_{j=1}^K \beta_j C_{i,j} + \gamma E_{US_i} + \varepsilon_i$$

where  $K$  is the number of themes, or topics, and  $E_{US_i}$  is the average business establishment entry rate of the US during governor  $i$ ’s term.  $E_{US_i}$  is used as a control for the state of the US economy (the US business cycle).

## 2.2 Data and variables construction

The raw data used in this study is a corpus of 596 SoSA delivered by US governors from 2001 to 2013. As Ferguson (2006) notes, the SoSA are used by governors "to lay out their vision of the problems facing their states as well as their proposed solutions." Thus, these speeches are potentially useful in assessing the priorities and goals of political leaders.

### 2.2.1 Latent Dirichlet Allocation: an intuitive exposition

We construct the set of topics (and their relative importance) using the Latent Dirichlet Allocation algorithm, also called topic modeling (Blei et al. (2003), Blei (2012)). This section describes the topic modeling framework used in this paper. Before applying the algorithm, it is customary in the literature to preprocess the text documents. Link words (such as: a, the, and, for...) and common English words are dropped. Longer words are truncated to their roots. For example, education, educated and educating, having the same root are all truncated to “educ.” The resulting data consists of a matrix,  $W$ , with 596 rows each containing the distribution of words contained in a particular speech. That is, we construct a matrix of words counts where each row represents

a speech and each column represents a word. Clearly, the number of words can range in the thousands, hence the need for dimension reduction methods aimed at collapsing the matrix of words into a matrix of themes (or topics, or concepts). The LDA algorithm is one such dimension reduction method. It is a hierarchical Bayesian matrix factorization method that seeks to reduce the word distributions into combinations of a smaller set of distributions, called topics or themes. Intuitively, we seek to decompose  $W$  as follows:

$$W_{D,V} \simeq \theta_{D,K} \phi_{K,V}$$

where  $\theta$  is a matrix of topic distribution over documents, and  $\phi$  is a matrix of word distributions over topics.  $D$  represents the number of documents (or speeches), and  $V$  is the vocabulary list, i.e. the list of unique words (think of each word as a variable). We postulate that the above approximation works well for a value of  $K$  much smaller than  $D$ . In this case,  $\phi$  contains a collection of words grouped into a small number ( $K$ ) of topics and  $\theta$  contains the percentage of each topic contained in each speech. Thus, the first column of  $\theta$  represents the "Topic 1" proportions in each of the documents. The first row of  $\phi$  represents the words' relative importance for Topic 1. By sorting the first row of  $\phi$  by decreasing order of words relative importance, it is hoped that the sorted list can provide a meaning to Topic 1. The remaining topics are interpreted likewise. It may help to establish a parallelism with Principal Component Analysis (PCA). Indeed, LDA can be understood as a Bayesian approach to PCA, applied to count data. In this case, the  $\theta$  matrix is the matrix of component, and the  $\phi$  matrix is the matrix of loadings. A major difference between PCA and LDA is that the components and the loadings values are interpreted as probability values. It is also helpful to think of each element of the  $\theta$  matrix as an index, and the rows of  $\phi$  as the contribution of each word to the construction of the index. To aid the understanding of topic modeling, the following examples of the LDA outputs are provided. Assuming  $K$  (the number of topics) is 2, the topics distribution of the first six documents is shown in Table 1, which is an excerpt of  $\theta_{596,2}$ .

Table 1: Example of topics distribution when  $K$ , the number of topics imposed, is 2.

	Topic.1	Topic.2
Alabama_2001_D_1.txt	0.75	0.25
Alabama_2002_D_2.txt	0.65	0.35
Alabama_2003_R_3.txt	0.26	0.74
Alabama_2004_R_4.txt	0.38	0.62
Alabama_2005_R_5.txt	0.50	0.50
Alabama_2006_R_6.txt	0.45	0.55

The first row of Table 1 shows that Topic 1 occupies about 75% of Alabama's SoSA of 2001. Topic 2 occupies about 25%.

Table 1 suggests that the SoSA of Alabama in 2001 and 2002 are similar (they are very high on Topic 1), and those of 2003 and 2004 are also similar (they are very high on Topic 2). These similarities and/or differences would have been very difficult to detect by looking at the words distributions. Are these differences suggestive of policies or agendas differences? Do these differences translates into differences in economic outcomes? Are these differences suggestive of policy differences between governors? 3.1 argues for a yes answer to these questions.

Table 2 gives the first few words in the (transpose of the)  $\phi$  matrix. Note that in this table, the words relative weight are replaced with the words themselves. Thus, for each column, the first word is the word with the highest weight. Examining these words may be useful in interpreting the topics in Table 1 .

Table 2: List of words ranked by their relative importance for their respective topics. The list is used to infer the meaning of the topic.

Topic 1	Topic 2
school	budget
educ	fund
work	govern
help	peopl
econom	million
children	work
famili	make
health	public
busi	propos
nation	servic
make	chang
creat	program
student	know
teach	spend
invest	come

It appears that Topic 1 is about education and that Topic 2 is about funding. Thus we might surmise that the governor’s priorities in 2001 and 2002 were education, whereas the governor’s priorities in 2003 and 2004 were budgetary issues. Interestingly, the speeches of 2001 and 2002 were of a Democrat governor, and those of 2003 and 2004 were of a Republican governor. With  $K = 2$ , we already see some differences between a Democrat and a Republican Alabama’s governors.

Next, by moving from two to four topics ( $K = 4$ ), that is collapsing the matrix of words into a matrix of four topics, it appears that the speeches of 2002 and 2003 are similar with respect to Topic.3, and those of 2002, and 2004 are similar with respect to Topic.2 (see Table 3 ). By increasing the number of topics, i.e. the level of detailed decomposition of the speeches, we allow for finer exposition of differences or similarities between speeches.

In sum, two lessons are to be drawn from the examples: (1) by setting the number of topics, we decide on the level of detail information we want; and (2), the topics proportions informs on the differences and/or similarities between speeches or between governors.

If we want to use the contents of the speeches (i.e the matrix of topics distributions) to study differences and/or similarities between governors, the choice of  $K$  matters for the level of variations we can capture. Thus, it is important to explain the choice of  $K$  (See Section 2.2.3).

Table 3: Example of topics distribution when  $K$ , the number of topics imposed, is 4.

	Topic.1	Topic.2	Topic.3	Topic.4
Alabama_2001_D_1.txt	0.07	0.08	0.21	0.64
Alabama_2002_D_2.txt	0.07	0.11	0.36	0.46
Alabama_2003_R_3.txt	0.46	0.07	0.37	0.11
Alabama_2004_R_4.txt	0.34	0.12	0.34	0.21
Alabama_2005_R_5.txt	0.23	0.15	0.32	0.30
Alabama_2006_R_6.txt	0.23	0.10	0.30	0.37

### 2.2.2 Latent Dirichlet Allocation (LDA)

LDA is a generative model that represent documents as being generated by a random mixture over latent variables called topics (Blei et al., 2003). A topic is defined as a distribution over words. For a given corpus of  $D$  documents each of length  $N_d$ , the generative process for LDA is defined as follows:

1. For each topic  $k$ , draw a distribution over words  $\phi_k \sim \text{Dirichlet}(\beta)$  with  $k = \{1, 2, \dots, K\}$
2. For each document  $d$ :
  - (a) Draw a vector of topic proportions  $\theta_d \sim \text{Dirichlet}(\alpha)$
  - (b) For each word  $i$ 
    - i. Draw a topic assignment  $z_{d,n} \sim \text{multinomial}(\theta_d)$  with  $z_{d,n} \in \{1, 2, \dots, K\}$
    - ii. Draw a word  $w_{d,v} \sim \text{multinomial}(\phi_{k=z_{d,n}})$  with  $w_{d,v} \in \{1, 2, \dots, V\}$

With the above generative process, it is easy to write down the joint likelihood of the observed and hidden variables. Markov Chain Monte Carlo (MCMC), and Variational Bayes methods are often used to estimate the values for  $\theta$  and  $\phi$ . Though the generative process can be impressive in the level of mathematics involved, it is still useful to see the model as a form of PCA.

### 2.2.3 Deciding on the number of topics $K$

We use a traditional model selection approach to select  $K$ . That is, we relied on the adjusted r-squared method to select the  $K$ , the number of topics for which the regression model yields the highest adjusted r-squared. In practice, we iterate the regression model through different values for  $K$  ( $K = 2, 3, \dots, 50$ ), and for different dependent variable (one, two, and three period leads). Note that for our regression model, we use future business entry rate as our dependent variable. Figure 1 shows the values of the adjusted r-squared in the y-axis, and the  $K$  values in the x-axis. Respectively, panel A, B, C, and D represent the changes in the values of the adjusted r-squared as  $K$  changes when the dependent variables are the one, two, three period leads, and the combined values of the adjusted r-squared of the three models.

In light of Figure 1,  $K = \{4, 7, 8\}$  seem to be all reasonable choices for  $K$ . Indeed, though there are values of  $K$  for which the adjusted r-squared are bigger than the one given by  $K = \{4, 7, 8\}$ ,

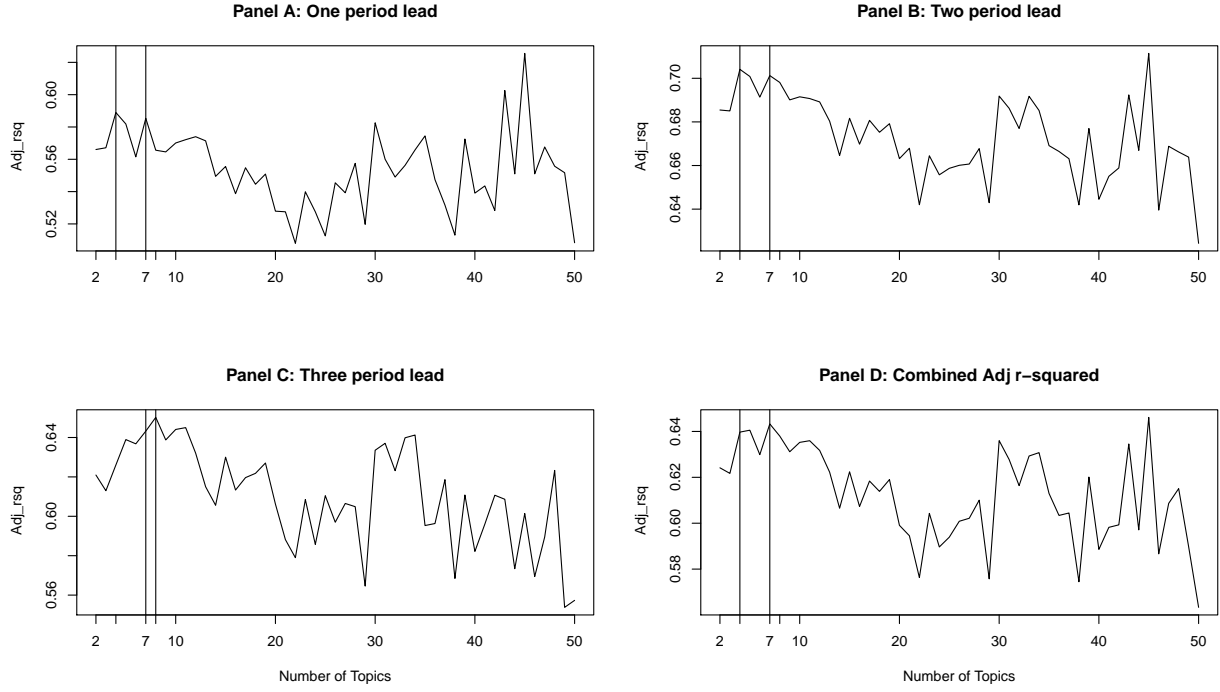


Figure 1: Values of Adjusted R-Squared as  $K$  changes.  $K = 4, 7,$  and  $8$  seem reasonable

those  $K$  are too large and would be difficult to interpret without further variables selection methods (such as LASSO, or subset selection). Consequently, we will consider  $K = 4$  as an optimum number of topics to work with. We will also use  $K = \{7, 8\}$  as a robustness check.

## 2.2.4 Consistency measure

By consistency we mean the persistence with which a governor addresses a particular issue in his/her SoSA. We believe that talking persistently about an issue is a sign of commitment to an agenda, which we term professed agenda after Grimmer (2010). To capture that idea of professed agenda, we use the log of the inverse of the coefficient of variation. Formally,

Let  $\theta_{i,j,l}$  be the relative share of topic  $j$ , ( $j = \{1, 2, \dots, K\}$ ), in governor  $i$ , ( $i = \{1, 2, \dots, N\}$ ) speech at year  $l$  ( $l = \{1, 2, \dots, L\}$ ). Then

$$\bar{\theta}_{i,j} = \frac{\sum_{l=1}^L \theta_{i,j,l}}{L},$$

gives an idea of the overall importance of topic  $j$  in governor  $i$  combined speeches.

$$s_{i,j} = \sqrt{\frac{\sum_{l=1}^L (\theta_{i,j,l} - \bar{\theta}_{i,j})^2}{L-1}},$$

gives the level of variations of topic  $j$  in governor  $i$  speeches; and

$$C_{i,j} = \log\left(\frac{\bar{\theta}_{i,j}}{s_{i,j}}\right)$$

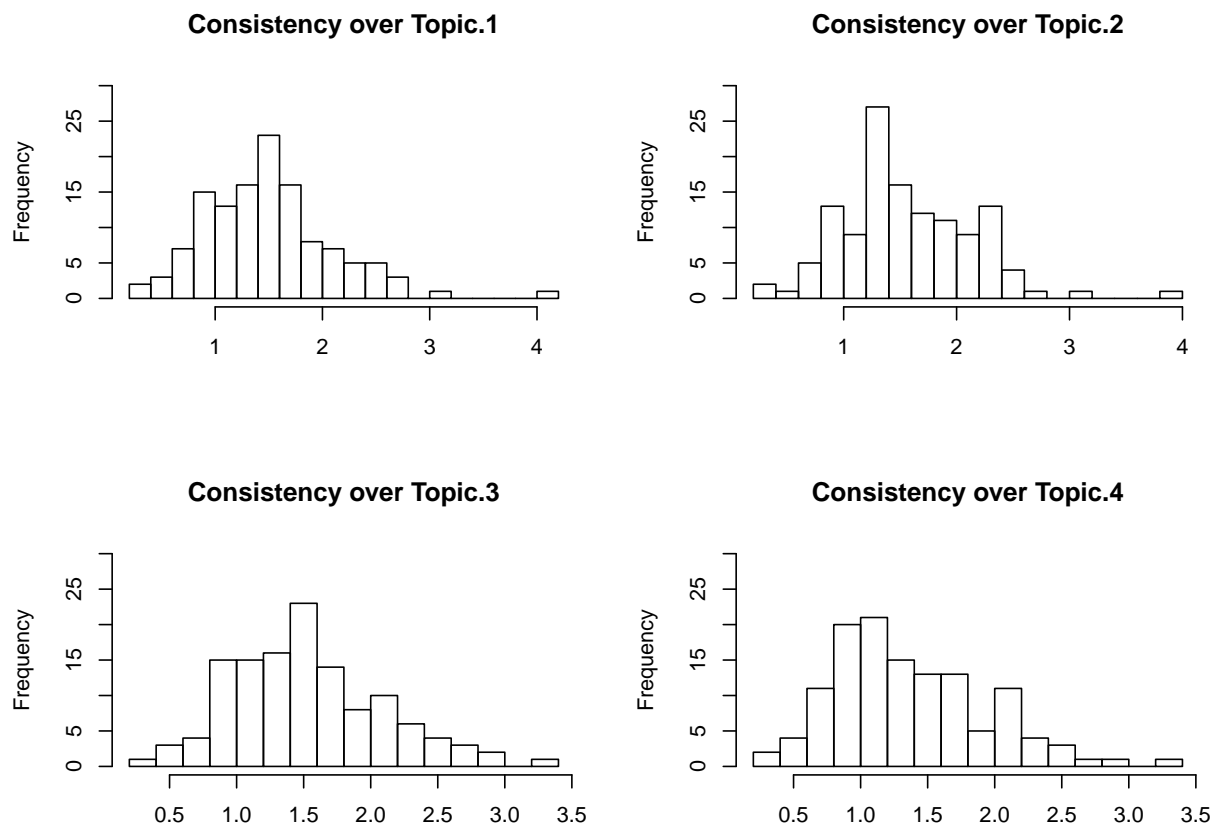


Figure 2: Distributions of the consistency measure for  $K = 4$  topics

is what we term consistency measure.

The intuition of our consistency measure is that consistency implies low variance. That idea is captured in the formula by having  $s_{i,j}$  in the denominator. However, a low variance alone is not enough to conclude that a particular topic or theme is important for a governor. Hence, we use the mean  $\bar{\theta}_{i,j}$  in the numerator. Consequently, our consistency measure is high when on average the governor talks a lot about the topic with low variations from year to year. We take the log of the ratio to temper the effect of potential outliers.

Figure 2 shows the distribution of the consistency measure by Topic. The figure shows that each variable has at most 2 slightly extreme values, and it is highly unlikely that our regression results are affected by high leverage observations. Put, differently, the regression results are robust.

### 2.2.5 The outcome variables: Establishment net entry rate

The dependent variable was collected from the US Census Bureau, Business Dynamics Statistics website. The annual business establishment entry rate (we are using the net entry rate) was



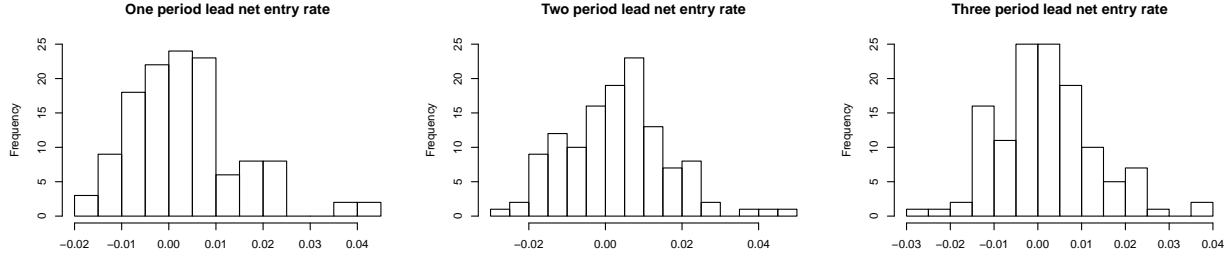


Figure 3: Distribution of establishment entry rate. From left to right, the distributions are for one, two, and three period leads.

computed using the following formula:

$$entry\_rate_{t,s} = \frac{estab_{t,s} - estab_{t-1,s}}{estab_{t,s}},$$

where:

- $entry\_rate_{t,s}$  is the business establishment net entry rate at year  $t$  in state  $s$ .
- $estab_{t,s}$  is the total number of business establishments at year  $t$  in state  $s$ .
- $estab_{t-1,s}$  is the total number of business establishments at year  $t-1$  in state  $s$ .

Similarly, we compute the US net entry rate as:

$$usa\_entry\_rate_{t,s} = \frac{usa\_estab_t - usa\_estab_{t-1}}{usa\_estab_t},$$

Because we postulate that if a governor action has to have an impact in the economy, that impact will be observed in the future. So we match 2001 SoSA with 2002 establishment net entry rate, which we refer to as one period lead. We call it two period lead if we match 2001 SoSA with 2003 net entry rate, and three period lead if we match 2001 SoSA with 2004 net entry rate.

The data are aggregated by governor term of four years. Therefore, for a governor of 2001 to 2004, the one period lead average net entry rate is computed as follows:

$$E_i = \frac{\sum_{t=2002}^{2005} NetEntryRate_t}{4}$$

Figure 3 shows that the outcome variables are fairly normally distributed, suggesting that it is appropriate to use OLS as an estimation method of the regression model.

## 3 Results

### 3.1 Regression results

#### 3.1.1 Regressing only on topics and us entry rate

We apply an OLS to the data set of  $N = 125$  observations, which refers to governors for whom we have three or four addresses. Table 4 shows some significant results.

Table 4: OLS with 1 period lead (1), OLS with 2 period lead (2), OLS with 3 period lead (3)

	<i>Dependent variable:</i>		
	entry_rate1 (1)	entry_rate2 (2)	entry_rate3 (3)
usa_rate1	1.000*** (0.081)		
usa_rate2		1.100*** (0.064)	
usa_rate3			1.000*** (0.074)
Topic.1	0.002* (0.001)	0.002* (0.001)	0.001 (0.001)
Topic.2	0.003** (0.001)	0.003*** (0.001)	0.003** (0.001)
Topic.3	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Topic.4	-0.001 (0.001)	-0.001 (0.001)	-0.0001 (0.001)
Constant	-0.007** (0.003)	-0.007** (0.003)	-0.007*** (0.003)
Observations	125	125	125
R <sup>2</sup>	0.610	0.720	0.640
Adjusted R <sup>2</sup>	0.590	0.700	0.630
Residual Std. Error (df = 119)	0.008	0.007	0.007
F Statistic (df = 5; 119)	37.000***	60.000***	43.000***

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Consistency on Topic.2 seems to be strongly correlated with establishment entry rate, whether we use the one, two, or three period lead, suggesting that our consistency measure of Topic.2 capture something meaningful in explaining the variations in the establishment entry rate variables. What is Topic.2 referring to? Clearly, Table 5 suggests that Topic.2 refers to an economic agenda.

Table 5: List of words, ranked by their relative importance for their respective topics. The list is used to infer the meaning of the topic.

Topic 1	Topic 2	Topic 3	Topic 4
fund	econom	peopl	school
budget	busi	govern	educ
million	work	make	children
increas	creat	work	health
propos	energi	know	teach
program	develop	come	student
servic	nation	just	famili
govern	futur	governor	help
provid	build	chang	make
dollar	help	like	care
revenu	invest	reform	high
system	compani	look	program
legisl	peopl	money	nation
educ	produc	good	colleg
depart	communiti	right	work

Our results suggest that the governor commitment to his/her economic agenda is positively and strongly associated with an increase in the number of business establishment in his/her state.

### 3.1.2 Adding party affiliation

Previous studies [cite?] on the role of leadership and economic growth have focused on comparing Democrats vs. Republicans. We use party affiliation as a control variable to check whether party affiliation mater. Table 6 suggests that it doesn't. Thus, it can be argued that being a Democrat or Republican doesn't matter in term of promoting business establishments. The governor business friendly agenda appears to matter more than his/her party affiliation.

### 3.1.3 Making sense of the results

Figure 4 shows the relationship between consistency on Topic.2 (or economic issues) and net entry rate of business establishments.

Clearly, there is a positive correlation between states' business establishment rate and the commitment level of the state governor to attract businesses. Our commitment measure captures what the governors profess to be doing. With respect to leadership studies, one major benefit of topic modeling is that we can refer to the leaders speeches to identifies what the governors profess to

Table 6: OLS with 1 period lead (1), OLS with 2 period lead (2), OLS with 3 period lead (3)

	<i>Dependent variable:</i>		
	entry_rate1	entry_rate2	entry_rate3
	(1)	(2)	(3)
usa_rate1	1.000*** (0.081)		
usa_rate2		1.100*** (0.065)	
usa_rate3			1.000*** (0.074)
partyR	0.002 (0.001)	0.001 (0.001)	0.0003 (0.001)
Topic.1	0.002* (0.001)	0.002* (0.001)	0.001 (0.001)
Topic.2	0.003*** (0.001)	0.004*** (0.001)	0.003** (0.001)
Topic.3	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Topic.4	-0.001 (0.001)	-0.001 (0.001)	0.00001 (0.001)
Constant	-0.008** (0.003)	-0.007** (0.003)	-0.007*** (0.003)
Observations	124	124	124
R <sup>2</sup>	0.610	0.720	0.640
Adjusted R <sup>2</sup>	0.590	0.700	0.620
Residual Std. Error (df = 117)	0.008	0.007	0.007
F Statistic (df = 6; 117)	31.000***	50.000***	35.000***

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

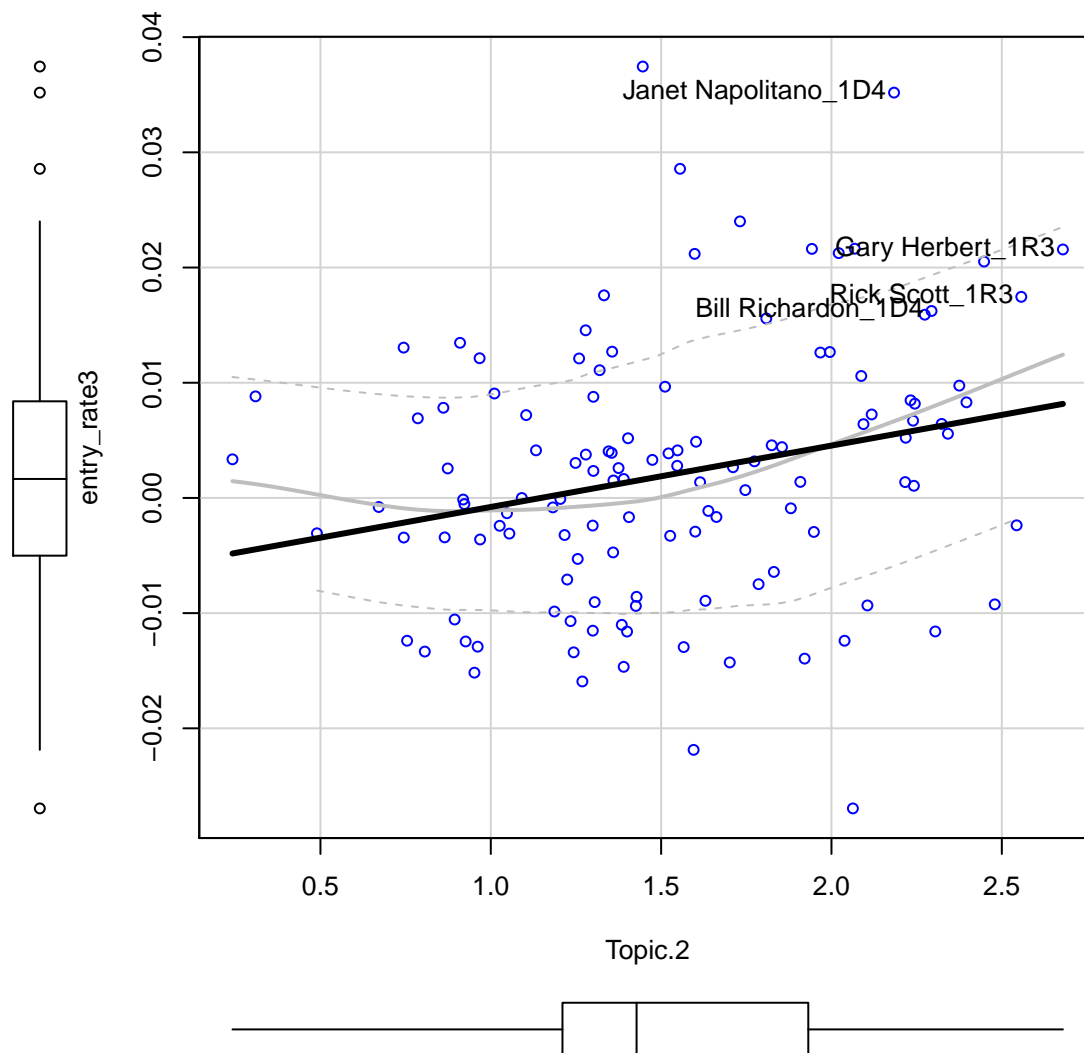


Figure 4: Scatter plot between topic.2 and entry rate. Topic.2 seems to refer to the governor economic agenda

be doing. The excerpts (Figure 5) highlight the content of four speeches by four “high achieving” governors, whom we identified on the scatterplot in Figure 4 (Janet Napolitano of Arizona, Gary Herbert of Utah, Rick Scott of Florida, and Bill Richardson of New Mexico). The excerpts shows policy actions, which are indeed expected to increase the number of establishments in the states if they are successful.

Figure 5: Excerpt of a few state’ State of State Addresses (1)

(a) Excerpt of the state of state address (Arizona 2003)

educate the workforces of tomorrow, and their research expands our horizons. Priority 2 Building the new Arizona **economy** Let's now turn to building the new Arizona **economy**. Although our unemployment rate is below the national average, we do not offer enough meaningful, high-paying, jobs for our people, whose per capita income ranks 37th among the states. It is time to coordinate our efforts to **develop** Arizona's promising tech sector. And it is time to take full advantage of our geographic proximity to establish Arizona as America's premier portal to trade with Latin America. **Develop** tech industries To achieve the **economic** renaissance I envision, our **economy** must be powered by **innovation**, and be driven by the entrepreneurs and tech-based **businesses** that will create the high-wage jobs and clean industries we seek. Three steps are key. First, our public and private sectors must speak with one **economic** voice. To do this, I will sign an executive order this week creating the Governor's Council on **Technology** and **Innovation**. It will focus on three areas: coordination of **technology** transfer from universities to the commercial sector, capital formation, and infrastructure **development**. Second, it is time to remove the single biggest obstacle to smooth **technology** transfer from our university campuses to the commercial sector. Arizona's constitution prohibits universities from forming or taking equity positions in commercial ventures, which slows down their efforts to convert research **innovations** into viable commercial applications. Competing states do not have this prohibition, and they enjoy greater success in luring tech start-ups. I will submit to you a ballot referendum to repeal this article of Arizona's constitution. Third, we must do a better job at attracting capital for small and growing high tech **businesses**. They need this capital to grow their enterprises and create more high-paying jobs for an educated workforce. I will work with the **business** community to attract more **development** capital to Arizona, particularly for new companies. International trade My administration will not limit its **economic development** to Arizona alone. Though many believe that our

(b) New Mexico 2003

and its entrepreneurs who play such a role in New Mexico's **economic** future. This administration is going to beat the bushes from coast to coast, from Europe to the Pacific Rim, seeking quality **companies** in need of good workers and a great environment into which to expand their **companies**. But, our **economic development** strategy doesn't depend on **industrial** recruitment alone. We must make the state of New Mexico a hospitable place for entrepreneurs to start and **grow** their **business** ideas. Whether from outside or from homegrown visionaries, ground-up **business development** must be encouraged and nurtured by state policy. I have already said I intend to spend at least 25 percent of my time working on **economic development**. We will undertake these efforts at many levels and with many concepts. **Growing** tourism and trade with Mexico is a theater of **development** with great promise particularly with our sister border state of Chihuahua. In cooperation with Chihuahua Gov. Patricio Martinez, I will work to build up the **business** and cultural bonds between our people. As New Mexicans learn more about Chihuahua, more will wish to visit and do **business** there. The same is true the other way to New Mexico. Here at home, we need to strengthen the teaching of **business** and entrepreneurial skills in our schools. Our young people must be taught the basics of **business** risk and reward so that more of them will take to improving their lives and building the **economy** of our state from within. Small **business** incubator programs in the population centers of the state must be strengthened and improved. Access to capital is critical, and because we lie so far from the money centers of the coasts, we remain below the radar of much of the venture capital market. The Legislature **invested** \$10 million in the New Mexico Small **Business Investment** Corp., but the money has languished in a bank account for lack of an implementation strategy. I will build that implementation strategy. I further propose that we **invest** up to \$200 million just 2 percent of the total in the state's permanent funds in New Mexico **businesses**. This will jump-start an entrepreneurial arm of New Mexico's **economy**. We will work with existing grass-roots **business** startup organizations such as Accion, Wesst Corp., the New Mexico Community **Development** Fund. With our new state **investment** officer, we have ensured the best expertise to manage and control the commitment of state venture funds. We will partner with private capital. While the primary purpose of our state's permanent funds must always be to provide revenue to state government, we must also **invest** them where fiscally prudent to create jobs and diversify our **economy**. By stepping up with cash, we will send a signal that New Mexico is serious about **business** and willing to put our money where our recruitment is. To facilitate all these ambitious **development** goals, I ask the Legislature for seed money: I would like to add \$15 million to the in-plant training fund, bringing it to \$20 million when combined with existing funds. \$3 million to fun a nonprofit corporation to recruit and market new **businesses** and jobs. We must tap the skills and leverage the efforts of everyone to **grow** the **economy**. \$9 million in a one-shot expenditure

Figure 6: Excerpt of a few state' State of State Addresses (2)

(a) Utah 2011

economic future secure. The third cornerstone essential to our return to prosperity is all about JOBS. My vision for economic development is that Utah will lead the nation as the best performing economy and be recognized as a premier global business destination. In Utah, we know, it is the private sector, not government, that creates jobs. And those jobs are being created through the expansion of homegrown Utah companies, as well as new companies relocating to our state. Some of the most recognized businesses in the world now call Utah home companies like Adobe, Proctor and Gamble, eBay, Litehouse Foods, Disney, Goldman Sachs, and the Royal Bank of Scotland, to name just a few. Additionally, local Utah businesses are expanding, like Petersen Inc, Nelson Laboratories, Lineagen, Merit Medical, Edwards Lifesciences, IMFlash, and Overstock.com. To accelerate this job creation across the state, we must focus on three key areas: First, we must increase access to capital, for our small and start up businesses. We must ensure that the Utah Fund of Funds, created by the Legislature three years ago, is focused on assisting UTAH companies. Second, we must expand our GLOBAL vision. Utah's export growth is the strongest in the nation. To ensure a continued focus on international business, I challenge Lew Cramer and other international business leaders to double Utah exports in the next five years. Third, I urge the Legislature to pass Senator Ralph Okerlund's Business Expansion and Retention bill to support companies throughout rural Utah. Utah has been recognized time and again as a pro-business state, including, for the first time in our state's history, a #1 ranking from Forbes as the "Best State for Business and Careers" in America. I am thrilled but not surprised we are the best place for business because we have the best people for business. However, the competition is getting tougher. My fellow governors across the country have all promised to improve their state economies. They are gunning for Utah's top spot for job growth. To stay ahead of the competition we must refine, distinguish, and promote our competitive advantages. One of those advantages is our unprecedented partnerships. I thank Senator Scott Jenkins for running legislation to create a Governor's Economic Development Coordinating Council. This council will ensure that the collective efforts of government and the business community are focused on jobs, jobs and more JOBS. This collaboration will be further enhanced by the co-location of many economic development

(b) Florida 2011

the corporate tax. These leaders, like me, share a positive view of Floridas economic potential. On behalf of the people of Florida, I want to thank all of you for your faith in Floridas future. I urge every member of the Legislature to join me in making job recruitment a daily task. I want to encourage each of you to become a Jobs Ambassador and direct new prospects to me, so we can work together to recruit potential job creators. Ask Florida business owners, What can we do to help you expand your business? Ask business leaders around the world, Why not move to Florida? Last July I submitted a detailed plan to the people of Florida to create 700,000 jobs over seven years. They reviewed the plan and voted to enact it. Last month, I delivered to you a budget that puts that plan into action and cuts taxes by \$2 billion. These tax cuts put money back in the hands of families and business owners who will grow private sector jobs. An important priority in our jobs budget is to consolidate governments economic development efforts into a single, highly focused agency. Working with our public-private partner, we will have the resources to be effective, and the flexibility to adapt to particularly promising opportunities. This agency will be headquartered two doors down from my office, and its work will never be far from my mind. I come to the job of Governor after a 35-year career in the private sector. I want to use that business experience on behalf of the people of Florida. Im asking this legislature and the people of Florida to give me the tools and hold me accountable for results. Our jobs budget makes sure government is held accountable for every spending decision. And by focusing on the core missions of government and only the core missions this budget will give Florida a competitive edge in attracting jobs. I know the members of this body have thoughtful, constructive modifications to our jobs budget. But we must not lose our focus or blunt our momentum. Business people in Florida and around the world are watching what we do in the weeks ahead. They can locate anywhere. They will be deciding whether to invest in Florida, based, in part, on our ability to work together to remove the obstacles to business success. I am convinced that putting this plan into action will put our state on the road to prosperity. On behalf of the millions of Floridians who are desperate for new jobs, I ask you to pass our jobs budget promptly. We also

### 3.2 Robustness check

A major challenge associated with topic modeling is how to decide on the value of  $K$ , the number of topics. We ran, iteratively, several regressions looping through different values of  $K$  in order to select the value of  $K$  for which the topics explain the most the dependent variable. Though Figure 1 suggests  $K = 4$  is reasonable,  $K = 7$  and  $K = 8$  are also reasonable. Hence the question whether the regression results would be different had we chosen a value of  $K$  other than 4? We run the

regression model assuming  $K = 7$  (Table 7 ), and  $K = 8$  (Table 8 ). Topic.1 seems to be strong in both cases. By looking at the words rankings in Table 9 and Table 10, it is clear that Topic.1 in these tables correspond to Topic.2 in Table 5. Thus, consistency on the economic agenda is strongly and positively associated with higher establishment entry rate.

Table 7: OLS with 1 period lead (1), OLS with 2 period lead (2), OLS with 3 period lead (3)

	<i>Dependent variable:</i>		
	entry_rate1	entry_rate2	entry_rate3
	(1)	(2)	(3)
usa_rate1	1.000*** (0.082)		
usa_rate2		1.000*** (0.065)	
usa_rate3			1.000*** (0.072)
Topic.1	0.002* (0.001)	0.002** (0.001)	0.002** (0.001)
Topic.2	0.0002 (0.001)	0.001 (0.001)	0.001 (0.001)
Topic.3	-0.0002 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Topic.4	0.003** (0.002)	0.003* (0.001)	0.002* (0.001)
Topic.5	-0.001 (0.001)	-0.001 (0.001)	-0.0004 (0.001)
Topic.6	0.002 (0.001)	0.002* (0.001)	0.002* (0.001)
Topic.7	-0.001 (0.001)	-0.001 (0.001)	-0.0002 (0.001)
Constant	-0.007** (0.003)	-0.007** (0.003)	-0.009*** (0.003)
Observations	125	125	125
R <sup>2</sup>	0.610	0.720	0.670
Adjusted R <sup>2</sup>	0.580	0.700	0.640
Residual Std. Error (df = 116)	0.008	0.007	0.007
F Statistic (df = 8; 116)	23.000***	37.000***	29.000***

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01



Table 8: OLS with 1 period lead (1), OLS with 2 period lead (2), OLS with 3 period lead (3)

	<i>Dependent variable:</i>		
	entry_rate1 (1)	entry_rate2 (2)	entry_rate3 (3)
usa_rate1	0.980*** (0.083)		
usa_rate2		1.000*** (0.066)	
usa_rate3			1.000*** (0.073)
Topic.1	0.001 (0.001)	0.002** (0.001)	0.003*** (0.001)
Topic.2	0.003 (0.002)	0.002 (0.002)	0.002 (0.001)
Topic.3	0.001 (0.001)	0.002 (0.001)	0.002* (0.001)
Topic.4	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Topic.5	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Topic.6	-0.001 (0.001)	-0.001 (0.001)	-0.0002 (0.001)
Topic.7	0.0002 (0.001)	0.00004 (0.001)	0.0004 (0.001)
Topic.8	-0.0001 (0.001)	-0.001 (0.001)	-0.002 (0.001)
Constant	-0.007* (0.004)	-0.006* (0.003)	-0.007** (0.003)
Observations	125	125	125
R <sup>2</sup>	0.600	0.720	0.680
Adjusted R <sup>2</sup>	0.570	0.700	0.650
Residual Std. Error (df = 115)	0.008	0.007	0.007
F Statistic (df = 9; 115)	19.000***	33.000***	27.000***

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 9: List of words ranked by their relative importance for their respective topics. The list is used to infer the meaning of the topic.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7
busi	nation	peopl	school	budget	fund	health
econom	futur	know	educ	govern	million	care
energi	famili	just	student	econom	program	famili
creat	econom	make	teach	spend	budget	work
work	live	come	children	reform	increas	help
develop	work	work	high	busi	propos	make
compani	challeng	like	colleg	system	provid	cost
invest	serv	look	learn	educ	addit	insur
build	protect	good	program	make	develop	children
produc	governor	money	system	billion	legisl	provid
peopl	opportun	take	nation	dollar	servic	afford
make	member	said	invest	public	support	protect
industri	face	right	succeed	servic	includ	invest
help	commit	give	better	fund	continu	safe
technolog	leader	chang	make	reduc	project	program

Table 10: List of words ranked by their relative importance for their respective topics. The list is used to infer the meaning of the topic.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
busi	school	fund	nation	peopl	budget	health	econom
energi	educ	million	famili	know	govern	care	work
creat	teach	budget	children	make	econom	famili	educ
econom	student	program	work	just	spend	make	futur
compani	children	increas	serv	come	reform	cost	communiti
work	high	propos	live	work	busi	work	opportun
peopl	colleg	provid	histor	like	system	help	develop
invest	program	addit	protect	look	billion	insur	resourc
produc	learn	revenu	governor	said	make	children	servic
build	help	project	challeng	believ	public	afford	govern
develop	system	feder	member	chang	educ	protect	continu
help	read	legisl	help	take	cost	provid	ensur
make	parent	dollar	leader	money	dollar	communiti	provid
nation	succeed	continu	like	governor	servic	program	commit
high	classroom	includ	public	place	save	invest	public

## 4 Discussion

This paper aimed at using what US governors say in their SoSA to capture what the governors profess to be their agenda, and further study the correlation between their agendas and business dynamics in their states. Ideally, we would measure what they do and correlate it with desired outcomes such as economic growth, educational achievement, health and more. Since, we are unable to measure what leaders do, this paper shows that we can rely on what they profess to be doing as a proxy for what they do. The paper strongly suggests that the consistency with which US governors pursue businesses has a positive impact in expanding business establishment in their states. Governors are salesmen or saleswomen of state agendas. Their commitment to their agendas should matter for their success.

The power of text analytics in general, and topic modeling in particular is that we can identify through the leaders words what they profess to be doing. For instance, Figure 7 tells us what actions (entrepreneurial policies) the governor of Oregon is undertaking. Topic modeling combined with traditional statistical methods (scatter plots and regressions) can show us some interesting patterns. By referring to key speeches in the collection of the governors speeches, we can identify plausible explanations of the observed patterns. In the current paper, it is clear that there is a positive relationship between governors economic agendas and business dynamics. The content of the speeches seems to suggest that the relationship is not fortuitous. The robustness check confirmed the strength of the results.

We purposefully refrained from interpreting the size of the coefficients for obvious reasons. The coefficients are sensitive to the number of topics  $K$  used for the analysis. Moreover, it hard to be precise with words. However, we believe that the sign of the coefficients are informative in term of exploratory data analysis.

Figure 7: Excerpt of the State of State Address: Oregon 2003

tolerance and social justice - that nurtures the talents of all our citizens. And today I'm here to tell you that we're putting in place an economic development plan that is comprehensive, forward-looking, balanced and workable. We're calling it GROW - the Governor's Recovery for Oregon Workers. We've handed out a detailed description of this plan. But I want to mention some of the highlights. Let me begin with the name. I've said for the last two years that only way to put Oregonians back to work is to grow the economy. Here is how I intend to do that. We're going to take every step we can - as fast as we can. First, we're going to take important administrative actions. I've already signed an Executive Order that will create an Office of Regulatory Streamlining. The office will ease burdensome regulations and shorten the time it takes to get a permit. The Order also says that in the next six months, we'll have a website for one-stop information about what kinds of licenses and permits are required for different business activities. I've signed a second Executive Order that directs state agencies to immediately take actions that will increase the availability of shovel-ready industrial land. This step is long overdue. Today Oregon competes against states where a company can pick from numerous 25-acre or larger industrial sites and get the necessary permits very quickly. The number of sites in Oregon that meet this market requirement is essentially zero. Also, administratively, we're going to redirect the focus of the Community Solutions Office. Its constituency will still be our rural cities and towns. But I want CSO to be an economic opportunity office that delivers on projects that create jobs. That's what Oregon families need right now. The next part of our economic development plan is to increase investments in business, infrastructure and workforce development. I already mentioned that we're changing the direction of the Economic and Community Development Department toward business retention and recruitment. This is not just a matter of rewriting the mission statement. That is not enough. We must give the Department the tools it needs to attract and retain jobs and businesses. That means supporting workforce development, building infrastructure, and marketing Oregon. We need to make long overdue investments in our infrastructure. Oregon's bridges and roads are in disrepair - and our economy is paying the price.

## 5 Conclusion

The paper set up to explore the relationship between the consistency with which US governors address certain issues in their state of the state addresses and the business dynamics in their states. The paper showed that on average the higher a governor addresses economic issues the higher the growth of business establishment in his/her state. The content of the economic theme suggests that the commitment to entrepreneurial policies do yield positive results. Text analytics is intrinsically challenging because it is based on words and people use words differently. Though this paper suggest a leadership positive effect, causal inference remains slippery. Consequently, a future research agenda may rely on legislation documents where the language is somewhat codified, making the comparison between governors or states more robust.

## References

- Alexopoulos, M. and Cohen, J. (2015). The power of print: Uncertainty shocks, markets, and the economy. *International Review of Economics & Finance*, (0):–.
- Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.
- Baker, S. R., Bloom, N., and C, S. J. D. (2013). Measuring economic policy uncertainty.
- Blei, D. M. (2012). Probabilistic topic models. *Commun. ACM*, 55(4):77–84.
- 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>.
- Bunten, D., Weiler, S., Thompson, E., and Zahran, S. (2015). Entrepreneurship, information, and growth. *Journal of regional science*, 55(4):560–584.
- Einav, L. and Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210).
- Ferguson, M. (2006). *The Executive Branch of State Government: People, Process, and Politics*. ABC-CLIO’s about state government. ABC-CLIO.
- Gennaioli, N., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2013). Human capital and regional development. *The Quarterly Journal of Economics*, 128(1):105–164.
- 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.

- Kerr, W. and Nanda, R. (2009). Financing constraints and entrepreneurship. Technical report, National Bureau of Economic Research.
- Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.
- 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.
- Romp, W. and De Haan, J. (2007). Public capital and economic growth: A critical survey. *Perspektiven der Wirtschaftspolitik*, 8(S1):6–52.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28.
- Winter, D. G. (2005). Things i've learned about personality from studying political leaders at a distance<sup>1</sup>. *Journal of Personality*, 73(3):557–584.