The Political Economy of Vermont's Anti-Fracking Movement

Alexandre R. Scarcioffolo*
Shishir Shakya †
Joshua C. Hall ‡

2018

Abstract

In 2012, Vermont became the first state in the United States to ban hydraulic fracturing for natural gas and oil production. We explain the political economy of the anti-fracking movement in the Vermont General Assembly. We implement elastic-net penalized binomial logit regressions to assess the best model to explain voting outcomes by members of the Vermont House of Representatives. We control for legislator-specific characteristics, median voter preferences, and special interests. Our results show that Democrats were more likely to vote for a fracking ban, while legislators representing districts with a higher percentage of working-age population and with a higher poverty rate were more likely to vote against anti-fracking. Campaign donations from the energy sector did not play a role. We discuss implications for median voter preferences regarding energy policy development in the US.

Keywords: Anti-fracking, Vermont, median voter, special interests
JEL Classification: D72, D73

*Alexandre R. Scarcioffolo, Department of Resource Economics and Management & Regional Research Institute (RRI), West Virginia University, Morgantown WV 26506, Morgantown WV 26506 E-mail: ar0044@mix.wvu.edu

†Shishir Shakya, College of Business and Economics & Regional Research Institute (RRI), West Virginia University, Morgantown WV 26506 E-mail: shishir.shakya@mail.wvu.edu.

‡Joshua C. Hall, College of Business and Economics & Regional Research Institute (RRI), West Virginia University, Morgantown WV 26506 E-mail: joshua.hall@mail.wvu.edu.

1 Introduction

Recently, the production of natural gas and oil from shale and other lower-permeability formations has become feasible due to horizontal drilling and hydraulic fracturing (fracking) technologies. These technologies have boosted unconventional natural gas production in the United States. As an effect, some Appalachian regions which were net importers of natural gas have become net exporters (Culver and Hong, 2016). Other than technological advancements, the Energy Policy Act of 2005 played a crucial role in stimulating the production of unconventional gas and oil. The act exempts fluids from fracking activities from the Underground Injection Control (UIC) provisions of the Safe Drinking Water Act (SDWA).

After the passage of the Energy Policy Act of 2005, policymakers in many regions became concerned about the possible negative effects of shale development. In some places, this culminated in local or state-level bans. A growing literature examines fracking regulations and bans from a political economy perspective (Rahm, 2011; Rabe and Borick, 2013; Boudet et al., 2014; Heikkila et al., 2014; Ritchie, 2014; Clarke et al., 2015, 2016; Evensen and Stedman, 2016; Hall et al., 2018). These studies look at locations that have known natural gas reserves and thus reflect both political and economic factors at play. In this paper, we look at Vermont, the first state in the United States to ban fracking. We do so because it allows us to better isolate the factors that drive expressive voting by legislators. Since Vermont has no known natural gas and oil reserves, voting to ban fracking was largely symbolic. This setting allows us to better understand the personal and constituent factors related to expressive voting on fracking. In doing so, our results inform the debate over the regulation or ban of fracking in other locations.

It is important to better understand expressive voting considerations from economic considerations given that from a narrow economic perspective, the exploration of lower-permeability formations of oil and natural gas has supported thousands of jobs together with the reduction of energy cost for taxpayers (Maniloff and Mastromonaco, 2017), and lowering natural gas prices throughout the US (Scarcioffolo and Etienne, 2018). However, some pundits suggest the activity also impact the region negatively. The main impacts of drilling unconventional oil and gas are twofold: environmental degradation and human health deterioration.[1] From the former, there are three central environmental concerns of a typical fracking process: first, the injection of chemicals into the ground which can contaminate the ground-water aquifers; second, methane seeping into the water supply system; and third, the manifestation of radioactive materials as result of the fracking process (Jenner and Lamadrid, 2013). There is a growing literature that measures the environmental impacts of the shale boom in the United States (Osborn et al., 2011; Vidic et al., 2013; Sovacool, 2014; Jackson et al., 2014; Silva et al., 2017). Most of these studies focus on water contamination related to gas/oil activity. Moreover, there are some indications that the fracturing process also contributes to increased seismicity and earthquakes activities in local areas (Sovacool, 2014).

Regarding health deterioration, in addition to the health impact of consuming contaminated water and being exposed to radioactive materials, the fracking process also plays a role concerning air pollution. Since most shale gas wells use diesel-powered pumps to inject water and depend heavily on trucks to deliver water to the site, there is evidence of higher levels of air pollution in regions with drilling activity (Sovacool, 2014). A study conducted in Pennsylvania shows that the number of hospitalizations for pneumonia among the elderly have increased in areas with fracking wells, indicating that the drilling process of gas and oil can contribute to air pollution, consequently affecting the well-being of the local population (Peng et al., 2018).

In response to possible environmental and health concerns, Vermont became the first state in the U.S. to ban fracking (Maur, 2015). On May 12, 2012, then Vermont Governor Peter Shumlin signed a bill that banned hydraulic fracturing wells from producing natural gas and oil. The ban is exclusively concerned with the environmental and public health impacts of

[1]According to Johnson et al. (1997) environmental degradation can be defined as any undesirable (destructive) disturbance or changes to the environment. The World Health Organization (WHO) defines human health as a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity. Then, human heath deterioration can be defined as any undesired changes on one's health (World Health Organization, 2006).

fracking.[2] More interesting, however, is that the state of Vermont does not have commercial production of natural gas or oil or known reserves.[3] The state's only natural gas utility receives its inputs from a small-capacity pipeline from Canada (US Energy Information Administration, 2018). Unlike Vermont, southwestern New York is located at two of the most important shale gas basins in the United States, the Marcellus and Utica formation (Wegener, 2014). Reflecting the possible adverse effects of drilling natural gas and oil, i.e., environmental and health deterioration, in 2010 some of New York's regions applied local zoning laws to ban fracking, followed by the state-level ban in December of 2014 (Podolny, 2013; Hall et al., 2018).

Using a median voter model, Hall et al. (2018) find that education levels, poverty rates, and veteran presence associate with the local ban. Since Vermont does not have proven shale reserves nor commercial production of fossil fuels an interesting question arises: what were the main determinants in banning fracking activity in a state that does not possess proven reserves of oil and natural gas? By following Hall et al. (2018), we isolate the ideology factors from political-economic factors that influence voting on fracking by exploring a situation where the vote seems to be purely expressive (Barton and Rodet, 2015; Lee and Murphy, 2017) and not directly related to economic benefits or costs. In doing so, we hope to better understand the sources and strength of expressive views regarding fracking.

In addition, our paper is the first of its kind in the median voter literature to utilize an elastic-net penalized binomial Logit regression. This is a machine learning approach, with the cross-validation method to evaluate the determinants that influence legislative votes. Our results suggest Democrats are likely to support the ban, while legislators who represent districts with higher poverty rate and population between 25 and 44 years are less likely to vote in favor of the ban. This later result highlights that even in a situation where there

[2] The Vermont legislature was given the power to revoke the moratorium on hydraulic fracturing for oil and natural gas if the fracking technique can be conducted without jeopardizing Vermont's environment (Department of Environmental Conservation, 2015).

[3] However, it cannot be ruled out the existence of these resources in the state (Department of Environmental Conservation, 2015).

is *zero* economic cost to banning fracking, politicians who represent areas more likely to benefit from fracking (high poverty and with more workers) are still likely to be in favor of fracking. This finding highlights the importance of looking at a situation where voting is purely expressive. Our results contribute to the energy policy literature by suggesting that better understanding how some groups of voters thought to be in favor of fracking purely for economic reasons may be in favor for non-economic reasons. Expressive NIMBYism - "Not In My Back Yard" (Fischel, 2001) by individuals in favor of fracking bans may encourage similar YIMBYism "Yes In My Back Yard" by other voters.

We proceed as follows. Section 2 explores the policy rationale on banning fracking. Section 3 presents our theoretical model, while Section 4 describes the data. Section 5 presents our empirical results and Section 6 concludes.

2 Policy rationale for banning fracking

There is a growing literature focused on better understating the determinants of energy policy related to fracking, since policymakers have to decide between the economic benefits that fracking activity brings to the state in the form of revenues, employment, and taxes, while addressing the environmental and health impacts from the oil and natural gas activity (Rahm, 2011; Rabe and Borick, 2013; Heikkila et al., 2014; Davis, 2017). It is clear that regions that have shale development perceive regulation as a mechanism to protect the environment from the possibility of negative externalities (Ritchie, 2014). However, what are the justifications for banning fracking in regions that do not produce oil nor natural gas? Possible explanations for why a state like Vermont banned fracking might be related but not limited to: public perception, geographic distance, and political-ideological statements. Political-ideological statements are described in the political science literature as "expressive voting".

A different explanation lies on the public view's perception of the fracking process (framing effects) as well as the geographic distance to areas under shale gas development (construal level theory) (Clarke et al., 2016, 2015; Evensen and Stedman, 2016). Under the construal level theory, areas geographically closer to shales gas development regions, i.e., southwestern Vermont, would be more likely to favor a ban on fracking since they can "experience" and "see" the impacts of the fracking activities on their neighbors (Clarke et al., 2016; Evensen and Stedman, 2016). Moreover, the way that the extraction of unconventional natural gas and oil is framed impacts the perception of the public view. As shown in Clarke et al. (2015), regions where the terminology "fracking" is more relevant than "shale gas development" is more likely to support the ban since the term "fracking" has a negative connotation.

Finally, voting behavior matters. As rational individuals, legislators take into consideration two rival accounts in order to decide their vote: instrumental and expressive accounts (Fiorina, 1976; Lomasky and Brennan, 1993; Brennan and Hamlin, 1998; Mueller and Günther, 2003). Instrumental voting is related to the electoral outcome that they believe is the best based on their own reading of the costs and benefits, while expressive account can be interpreted as an act of expressing opinion on what that outcome should be.[4] The Vermont case is interesting since a ban on fracking has no immediate cost or benefit, since Vermont has zero known natural gas and oil reserves. To the extent that legislator's follow the wishes of their constituents, their vote might reflect their constituents disapproval or approval of shale development in their region (Ritchie, 2014). As explained in Hall et al. (2018), when voting on a single issue such as whether banning fracking activities in a region or not, the majority rule voting system will select the outcome related to the person whose preference lies in the middle of the distribution of voter preferences. In doing so, the increasing support of a policy by communities enlarges the likelihood of the policy's adoption. Therefore, we can use the median voter model to analyze expressive voting in the case of banning fracking activities in Vermont. Several papers have utilized the median voter model to empirically analyze different political economy decision (Jones and Dunlap, 1992; Hall

[4]Please see Mueller and Günther (2003) for a further discussion.

and Schiefelbein, 2011; Boudet et al., 2014; Neto et al., 2016).

3 Theoretical Model

This section provides a theoretical model of how a utility maximizing individual legislator votes based upon her characteristics and the voter's characteristics which she represents. The researcher can access or collect data of the individual characteristics of the legislator, the voter's characteristics and the voting outcome (either for or against) hence, these are observable to a researcher. However, a researcher cannot access data on how the legislator weighs her costs and benefits of voting, thus these variables are unobservable. Based upon a few statistical assumptions, we provide a theoretical argument to link the unobservable – the net-benefit of the legislator or the latent regression – on how to estimate the probability of voting in favor for fracking bans with the observable variables of legislator's characteristics and the voter's characteristics which she represents.

Our model of legislative voting begins with an individual legislator. The conditional indirect utility function for a legislator g who has selected a voting option $y = \{0, 1\}$ can be written as:

$$U_{g,y} = V_{g,y}\left(x_{g,y}, \phi_{g,y}\right) + \varepsilon_{g,y} \tag{1}$$

where, $x_{g,y}$ are the observable characteristics of a legislator who selected a voting option $y = \{0, 1\}$ and $\phi_{g,y}$ are the median voter's observable characteristics in that legislator's district. This conditional indirect utility function consists of two part: deterministic $(V_{g,y}(\cdot))$ and random error part $(\varepsilon_{g,y})$. Although, the utility that a legislature receive voting 0 or 1 is unobserved. For simplicity sake, we drop the index g. We can define the unobserved utility as y_h^*. The unobserved utility that a given legislator derives from voting "Nay" can be represented as:

$$y_0^* = \mathbf{x'}_0\beta_0 + \varepsilon_0 \tag{2}$$

where, x'_0 includes two components. First, the observable characteristics of the legislator such as party affiliation, incumbent status, education, and political donations (Hall and Shultz, 2016). Second, the median voter's observable characteristics that the legislature represents like poverty rate, unemployment rate, working aged population, education, income. Similarly, the utility for voting "Yea" can be expressed as:

$$y_1^* = \mathbf{x'}_1\beta_1 + \varepsilon_1 \tag{3}$$

Voting also entails costs that may be related to the individual legislator. The real cost is also unobserved and can be defined as:

$$C^* = \mathbf{z}'\alpha + u \tag{4}$$

Any legislature will weigh his option to vote either "Nay" or "Yea" if the benefit of voting, i.e $y_1^* - y_0^*$ is greater than the cost C^*. The net benefit of voting is:

$$y^* = y_1^* - y_0^* - C^* = \mathbf{x'}_1\beta_1 - \mathbf{x'}_0\beta_0 - \mathbf{z}'\alpha + (\varepsilon_1 - \varepsilon_0 - u) = \mathbf{x}'\beta + \varepsilon \tag{5}$$

The net-benefit of the legislator is also not observable but the legislator's vote itself is observable. Following the approach of Greene (2003), the $y^* = \mathbf{x}'\beta + \varepsilon$ is the latent regression. We observe that $y = 1$ for a legislator when she votes "Yea" and $y = 0$ when a legislator votes "Nay". Therefore, our observation will be:

$$y = \{ 1 \quad if \quad y^* > 0 \quad or \quad 0 \quad if \quad y^* \leq 0 \tag{6}$$

Let a be a nonzero threshold[5] and α be an unknown constant term. $\mathbf{x}$ and β contain the

[5] The $\mathbf{x}'\beta$ is the index function and ε demands an assumption of its variance. Say, ε is scaled by an unrestricted parameter σ^2 then the latent regression is $y^* = \mathbf{x}'\beta + \sigma\varepsilon$ and $y^*/\sigma = \mathbf{x}'(\beta/\sigma) + \varepsilon$ represents the same model. Hence, the assumption of variance is an innocent normalization. The observed data remains unchanged; y is still 0 or 1, depending only on the sign of y^*not on its scale. This means that there is no information about σ in the sample data so σ cannot be estimated. The parameter vector β in this model

rest of the index not including the constant term. Then, the probability that a legislator vote "Yea" ($y = 1$) is:

$$P(y^* > a|\mathbf{x}) = P(\alpha + \mathbf{x}'\beta + \varepsilon > a|\mathbf{x}) = P((\alpha - a) + \mathbf{x}'\beta + \varepsilon > 0|\mathbf{x}) \tag{7}$$

Since, the $(\alpha - a)$ is an unknown parameter, the model has a constant term which remains for any choice of threshold α. If we normalize the constant to zero, then

$$P(y^* > 0|\mathbf{x}) = P(\varepsilon > -\mathbf{x}'\beta|\mathbf{x}) = P(\varepsilon < \mathbf{x}'\beta|\mathbf{x}) \tag{8}$$

The only assumption to be made is for the choice of distribution of ε. For any symmetrical cumulative distribution F, the probability of a legislator voting "Yea" is $P(y^* > 0|\mathbf{x}) = P(\varepsilon < \mathbf{x}'\beta|\mathbf{x}) = F(\mathbf{x}', \beta)$ and the probability of legislator voting "Nay" is:

$$P(y = 1|\mathbf{x}) = P(y^* > 0|\mathbf{x}) = P(\varepsilon < \mathbf{x}'\beta|\mathbf{x}) = 1 - F(\mathbf{x}', \beta) \tag{9}$$

The assumption of cumulative distribution largely depends upon assumption imposed in the ε. If the error terms are assumed to be *i.i.d.* with the extreme value distribution, then it assumed to be a logit model. Suppose the legislatures make the same assumptions subjectively, the voting to "Yea" probability would be in logit form as:

$$P(Y = 1|\mathbf{x}) = \frac{\exp(\mathbf{x}'\beta)}{1 + \exp(\mathbf{x}'\beta)} = \Lambda(\mathbf{x}'\beta) \tag{10}$$

4 Data

The dependent variable, vote by legislators (VOTE), was obtained from the Vermont General Assembly's website. If the legislator voted "Yea", the variable VOTE assumes the value

is only "identified up to scale." The assumption of zero for the threshold is likewise innocent if the model contains a constant term (and not if it does not) (Greene, 2003).

one, zero otherwise. Our independent variables represent legislator, median voter, and special interest factors (Congleton and Bennett, 1995). We web-scraped legislator's education and birthplace from the votesmart.org. As investigated by O'Roark and Wood (2011), the academic background of politicians sometimes matters for voting, as they find that members of Congress with undergraduate degrees in economics were less likely to vote for minimum wage legislation. If a legislator graduated with a bachelor's degree or higher in economics the variable (ECONOMICS) equals one, and zero otherwise. Similarly, we consider another dummy variable to represent whether the legislator majored in any natural resource field in college or not, which the variable (ENVIRONMENT) takes the value of one when it is true. We create another dummy variable (SUBJECT) that represents whether the variable ECONOMICS or ENVIRONMENT is equal to one.

To account the possible effect of being born in Vermont, the variable (BIRTHPLACE) assumes the value one when the legislator was born in Vermont, zero otherwise. This dummy variable sheds light on the intrinsic aspect of the local populations that are associated with a longer-term commitment to the state than non-locals. As presented in Hall and Shultz (2016), donation from the energy sector might influence voting behavior. Hence we consider the variable (%DONATION_ENNR_PCT) to represent the percentage of total donation that a legislature receives from contributors for the Energy and Natural Resource Sector, while the variable (DONATIONPER) represents the total amount of donations from the Energy and Natural Resource Sector. We also consider the total amount of donation that the legislator received (DONATION). Politicians that receive donations on the energy and natural resource sector are more likely to vote "Yea" since the energy matrix of the state is primarily renewable, and a possible, but unlikely, shale gas development in the state might increase competition to generate electricity since natural gas is more economical and attractive fuel for electricity generation due to its low price and high power capacity (US Energy Information Administration, 2017). We also collect legislator characteristics variables: (INCUMBENT) and (DEMOCRATS). The variable (INCUMBENT) assumes the value one if the legislator was an incumbent in the previous election, zero otherwise. Similarly, if the legislator is a Democrat, the variable (DEMOCRATS) assumes the value one, zero otherwise. We download the variables %DONATION_ENNR_PCT, DONATION, ENERGY, DONATIONPER, INCUMBENT and DEMOCRATS from followthemoney.org.

Additionally, we include variables to define the median voter of each district.[6] We gather household income (HHINC), the poverty rate (%POVERTY), the poverty rate under 18 years old (%POVU18), the poverty rate from age 5-17 years (%POVU517), and the percentage of the population that attend college (%SOMECOLLEGE) from Small Area Income and Poverty Estimates (SAIPE). The unemployment rate (%UNRATE) we obtain from the Bureau of Labor Statistics (BLS). The percentage of the population that has no high school degree (%NOSCHOOL), the percentage of the population that have high school degree (%HIGHSCHOOL), the percentage of the population that have a college degree (%COLLEGE) from the United States Department of Agriculture (USDA). Finally, from the American Community Survey (ACS) we obtained total population (POP), the percentage of the population above 65 years old (%POPA65), the percentage of the population under 18 years old (%POPU18), the percentage of the population between 18 and 24 years old (%POPU1824), the percentage of the population between 25 and 44 years old (%POPU2544), the percentage of the population between 45 and 64 years old (%POPU4565), the dependency ratio (%DEPENDENCY), males per 100 females for all ages (MP100F), males per 100 females over 18 years (MP100F18), the population between 25 and 44 years (%WORKFORCE), and the median age of the population median age (MAGE).

From the total of 180 legislators in Vermont, of which 150 belong to the House while the other 30 belong to the Senate, we only have data to consider the 167 legislators that participated in the voting process.[7] Table 1 shows the descriptive statistics of the data used

[6] It is important to clarify that due to the lack of data for each voting district, we match each legislator to a county. For districts that are located within two or more counties, we consider the average of the counties.

[7] Six legislators filled vacancy positions during the 2010-2012 term. All those six legislators were elected in the 2012 election. Due to the lack of data from them in the 2010 election, we considered the 2012 election data instead.

to understand the dynamics of legislator voting.

5 Econometric model and results

Under the median voter model, several possible explanatory variables might have influenced the legislator's decision to ban fracking in Vermont. Under such paucity of information, considering all possible variables would result in overfitting the model and consequently generate poor predictions (type-I error). On the other hand, not considering essential variables to explain the voting outcome would generate a biased estimator due to omitted-variable bias (type-II error). The type-I error leads to loss of efficiency whereas the type-II error leads to specification error. Furthermore, it is computationally infeasible to select the best regression specification by running all the possible combination regressions, in which in our case would be $2^{29} = 536,870,912$.

Given that we have to efficiently select the best logit model while addressing type-I and type-II error, we deal with high-dimensional data by using data reduction methods. When the likelihood function follows a normal distribution, the simplified method for dimension reduction in linear models is the principal components analysis (PCA). The PCA creates principal components using linear combinations of the original data set to reduce the number of variables in the model. However, this methodology suffers from a lack of interpretability of the principal component coefficients since they are a linear combination of several variables. Another alternative method of dimension reduction is to run regressions on the subset of explanatory variables, i.e., a variable selection approach. However, to decide the best model requires estimation of all possible combinations, totaling of 2^k models, where k is the number of variables (Nowak and Smith, 2017). Hence, applying variable selection is computationally infeasible.

A more realistic approach is to consider variable selection with penalized regression (Nowak and Smith, 2017). The penalized regression is a simple but powerful concept and is an essential development in modeling big data, one of the most popular statistical innovations of the past two decades (Hindman, 2015). The penalized regression is equivalent to performing least squares, just with added constraints on the coefficients or a penalty function to simultaneously generate a parsimonious model, allowing for coefficient estimation along with the predictive accuracy (Hindman, 2015). Three main models use this methodology: least absolute shrinkage selection operator (LASSO) (Tibshirani, 1996; Zou, 2006), ridge regression (Tikhinov regularization), and elastic-net (Zou and Hastie, 2005). The LASSO methodology forces the sum of the absolute value of the regression coefficients to be less than a certain threshold, forcing some coefficients to be set to zero. Similarly, the ridge regression forces the sum of the squares of the coefficients to be less than a fixed threshold; however, it does not set any of them to zero. Finally, elastic-net is a combination of both LASSO and ridge regression. Therefore, the penalized regression like LASSO, ridge or elastic-net (i) screens for important variables, (ii) provides easily interpreted coefficients and model, and (iii) performs well in out-of-sample testing, three features that are important in modeling the legislator voting decision.

The penalized regression approach has been widely employed in different fields such as estimating demand for energy (Aristondo and Onaindia, 2018; Miao et al., 2017; Raihanian Mashhadi and Behdad, 2018; Sardaro et al., 2018; Ziel and Weron, 2018), forecasting stock market performance (Chinco et al., 2017), forecasting movie success (Lehrer and Xie, 2016, 2018), effectiveness of workplace wellness programs (Jones et al., 2018), performance of energy efficiency programs at school (Burlig et al., 2017), measuring polarization in the congressional speech (Gentzkow et al., 2016), and predicting the recession on economy using fertility as leading indicator (Buckles et al., 2018).

We implement the elastic-net penalized binomial logit regression. We use the Hastie and Qian (2014)'s algorithm called the penalized binomial logit regression to select the best model and 10-fold cross-validation to generate reproducibility and uncertainty. The objective function for penalized logistic regression uses the negative binomial log-likelihood, and it is given as:

$$(\beta_0, \beta) \in R^{p+1} \min - \left[\frac{1}{N} \sum_{i=1}^{N} y_i \left(\beta_0 + {x'}_i \beta \right) - \log \left(1 + e^{\beta_0 + {x'}_i \beta} \right) \right] + \lambda \left[(1 - \alpha) \left\| \beta \right\|_2^2 / 2 \; + \alpha \|\beta\|_1 \right] \tag{11}$$

Hastie and Qian (2014)'s algorithm uses a quadratic approximation to the log-likelihood, and then coordinate descent on the resulting penalized weighted least-squares problem to solves Equation (11) over a grid of value of λ covering the entire range. The elastic-net penalty is controlled by α, and bridges the gap between lasso ($\alpha = 1$, the default) and ridge ($\alpha = 0$). The tuning parameter λ controls the overall strength of the penalty. Logistic regressions often degenerates when numbers of variables (p) are larger than numbers of observations (N), i.e. $p > N$, and exhibits erratic behavior even when N is close to p. The elastic-net penalty alleviates these issues, and regularizes and selects variables as well (Hastie and Qian, 2014). We run the elastic-net penalized binomial logit regression method for 1000 times implementing bootstrapping.

Primarily, we are searching for the "best" model. Table 3 presents seven models that are chosen as the best fit models[8]. The first column shows the list of variables that are selected while the second column shows the frequency of the variable selected while we run 1000 bootstrapped models. For example, while we run 1000 different bootstrapped models, the variables DONATION, BIRTHPLACE, DEMOCRATIC are selected for 1000 times while the variable %POPU2544 is selected for 969 times out of 1000 times, and ENVIRONMENT, %POPU1824, and %POPA65 were never chosen. The remaining columns show the models and how often each model is selected. Model V8 (model with 8 variables) is selected 268 times by elastic-net penalized binomial logit regression followed by V7, V12, V11, V24, V6, and V16A.

From this econometric methodology, we can infer how control variables influence the

[8]Full results available upon request.

probability of the legislators to vote “Yea” or “Nay”. However, before inference, we first perform variable selection technique and model validity tests to provide statistical justification of our inference from the list of variables selected from the literature and intuition informed by the literature.

5.1 Model Estimation

We identify five of the top seven models in our model estimation. (We exclude V24 in favor of V6 out of variable parsimony). We do not presume the validity of these models yet. To test the validity of models, the estimated coefficients must be statistically non-zero, and the model should accurately classify voting behavior out-of-sample. To test model validity, we perform 1000 bootstrapping exercises. Table 3 presents the estimates of these five different models with a) the coefficient estimates, b) standard errors, c) average marginal effects, d) average out-of-sample accuracy and its standard deviation, e) Akaike Information Criterion (AIC), f) McFadden's pseudo-R squared.

The out-of-sample analysis provides evidence on of the accuracy of the trained model to correctly classify “Nay” and “Yea” on the test data. We look for accuracy tests that are greater than the No Information Rates (NIR), in our case 77% (the percentage “Yea” votes), to justify model validity relative to the many voting scenarios that nature could have provided. All models presented in Table 3 have an accuracy greater than 0.77, with the models V8, V9 and V6 having the highest accuracy indicator. In respect to the AIC information, the model V6 presents the lowest information criteria. Hence the V6 model is the preferred model to analyze voting behavior. Another way to decide the best model is to choose the model with the least difference between the McFadden's pseudo-R squared to the adjust McFadden's pseudo-R squared. By these criteria, model V6 is also preferred.[9]

The variables DONATION, DEMOCRATIC, %POPU2544 and %POVERTY in the model

[9]The R-squared tend to reach toward one if we add more variables to the model, but the adjusted R-squared penalizes for the addition of more variables. Thus if the difference between R-squared and adjusted R-squared is close we feel it represents the best fit.

V6 are significant to explain the voting outcome. A Democrat legislator was, on average, 29% more likely to vote "Yea" on the fracking ban. This estimate is statistically significant at the 1% level. Since the Democrat party proposed the Act, this result is not surprising. Interestingly, being a native legislator decreases the likelihood of voting "Yea", which appears contrary to our expectation, however, this result is not statistically significant. We find positive but small effects of campaign donations on the likelihood of voting "Yea". The existence of this relationship cannot be rejected. Unlike Hall and Shultz (2016) who find donations from the energy sector matter for voting on the Keystone XL pipeline, we find no relationship between campaign contributions from the energy sector and legislator votes. We find a positive, but not statistically significant relationship between ECONOMICS training "Yea" voting.[10]

Considering the median voter variables, we find evidence that the %POPU2544 and %POVERTY are significant and negatively relate to the likelihood of voting "Yea". Legislators that represent districts with higher poverty rate were more likely to vote "Nay" on the fracking ban. Such a result is consistent with the environmental Kuznets curve hypothesis (Yandle et al., 2002), which postulates rich regions tend to be more concerned with environmental quality. Similarly, a greater population between 25 and 44 years, which are more likely to be of prime working age, is associated with voting "Nay" on banning fracking. This result highlights the benefit of studying the fracking ban in a state with zero known natural gas reserves. Even in a state with zero economic cost to banning fracking, constituent groups typically thought to bear the brunt of fracking bans for economic reasons are still against fracking bans even though there is no economic cost to them.

[10] As pointed out by an astute referee, we do not control for overall legislator education, which may explain this result.

6 Conclusion and policy implications

Our results indicate that being a Democrat impacted positively voting for anti-fracking. In that respect, our results are not surprising and match the literature on fracking bans (Boudet et al., 2014; Davis, 2017). Clarke et al. (2016) shows that political ideology is weighted heavier the more geographically removed the population is from the fracking activity. Since Vermont does not have commercial production of natural gas nor oil, voting on fracking should reflect the expressive, not instrumental views, of voters. We find that voter characteristics based on the economy still matter: a higher poverty rate and the population between 25 and 44 years explain votes against fracking. Legislators that represent districts with a higher poverty rate and/or with working population age are less likely to vote in favor of the ban; a similar result was found in Jones and Dunlap (1992). This highlights the importance of studying the Vermont case, since there are no known natural gas or oil reserves in the state, all constituent interests should be expressive, not instrumental.

Our results contribute to the literature on energy policy by showing that the increase in petitions ban fracking might be explained by expressive voting, especially regions that do not have good shale gas development plays, such as Florida and Delaware. Those regions are more concerned with the expansion of the activity even though they are not (and are unlikely to be) an active player in the fracking industry. Hence, citizen views about what is good or bad for the environment or health play a very important role on both sides of the debate. Areas with more poverty or larger working age populations were more likely to be against a ban, even though there was zero cost to them.

References

Aristondo, O. and Onaindia, E. (2018). Counting energy poverty in Spain between 2004 and 2015. *Energy Policy*, 113(November 2017):420–429.

Barton, J. and Rodet, C. (2015). Are political statements only expressive? an experiment. *Journal of Economic Behavior & Organization*, 116:174–186.

Boudet, H., Clarke, C., Bugden, D., Maibach, E., Roser-Renouf, C., and Leiserowitz, A. (2014). "Fracking" controversy and communication: Using national survey data to understand public perceptions of hydraulic fracturing. *Energy Policy*, 65:57–67.

Brennan, G. and Hamlin, A. (1998). Expressive voting and electoral equilibrium. *Public choice*, 95(1-2):149–175.

Buckles, K., Hungerman, D., and Lugauer, S. (2018). Is fertility a leading economic indicator? Technical report, National Bureau of Economic Research.

Burlig, F., Knittel, C., Rapson, D., Reguant, M., and Wolfram, C. (2017). Machine learning from schools about energy efficiency. Working Paper 23908, National Bureau of Economic Research.

Chinco, A. M., Clark-Joseph, A. D., and Ye, M. (2017). Sparse signals in the cross-section of returns. Working Paper 23933, National Bureau of Economic Research.

Clarke, C. E., Bugden, D., Hart, P. S., Stedman, R. C., Jacquet, J. B., Evensen, D. T., and Boudet, H. S. (2016). How geographic distance and political ideology interact to influence public perception of unconventional oil/natural gas development. *Energy Policy*, 97:301–309.

Clarke, C. E., Hart, P. S., Schuldt, J. P., Evensen, D. T., Boudet, H. S., Jacquet, J. B., and Stedman, R. C. (2015). Public opinion on energy development: The interplay of issue framing, top-of-mind associations, and political ideology. *Energy Policy*, 81:131–140.

Congleton, R. D. and Bennett, R. W. (1995). On the political economy of state highway expenditures: Some evidence of the relative performance of alternative public choice models. *Public choice*, 84(1):1–24.

Culver, W. J. and Hong, M. (2016). Coal's decline: Driven by policy or technology? *The Electricity Journal*, 29(7):50–61.

Davis, C. (2017). Fracking and environmental protection: An analysis of US state policies. *The Extractive Industries and Society*, 4(1):63–68.

Department of Environmental Conservation (2015). *A report on the regulation and safety of hydraulic fracturing for oil or natural gas recovery.* Vermont Agency of Natural Resources, Montpelier, VT.

Evensen, D. and Stedman, R. (2016). Scale matters: Variation in perceptions of shale gas development across national, state, and local levels. *Energy research & social science*, 20:14–21.

Fiorina, M. P. (1976). The voting decision: instrumental and expressive aspects. *The Journal of Politics*, 38(2):390–413.

Fischel, W. A. (2001). Why are there nimbys? *Land economics*, 77(1):144–152.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2016). Measuring group differences in high-dimensional choices: Method and application to congressional speech. Working Paper 22423, National Bureau of Economic Research.

Greene, W. H. (2003). *Econometric Analysis.* Pearson Education, New York.

Hall, J. C. and Schiefelbein, J. (2011). The political economy of medical marijuana laws. *Atlantic Economic Journal*, 39(2):197.

Hall, J. C. and Shultz, C. (2016). Determinants of voting behaviour on the Keystone XL pipeline. *Applied Economics Letters*, 23(7):498–500.

Hall, J. C., Shultz, C., and Stephenson, E. F. (2018). The political economy of local fracking bans. *Journal of Economics and Finance*, 42(2):397–408.

Hastie, T. and Qian, J. (2014). Glmnet Vignette.

Heikkila, T., Pierce, J. J., Gallaher, S., Kagan, J., Crow, D. A., and Weible, C. M. (2014). Understanding a period of policy change: The case of hydraulic fracturing disclosure policy in colorado. *Review of Policy Research*, 31(2):65–87.

Hindman, M. (2015). Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences. *Annals of the American Academy of Political and Social Science*, 659(1):48–62.

Jackson, R. B., Vengosh, A., Carey, J. W., Davies, R. J., Darrah, T. H., O'sullivan, F., and Pétron, G. (2014). The environmental costs and benefits of fracking. *Annual Review of Environment and Resources*, 39:327–362.

Jenner, S. and Lamadrid, A. J. (2013). Shale gas vs. coal: Policy implications from environmental impact comparisons of shale gas, conventional gas, and coal on air, water, and land in the United States. *Energy Policy*, 53:442–453.

Johnson, D. L., Ambrose, S. H., Bassett, T. J., Bowen, M. L., Crummey, D. E., Isaacson, J. S., Johnson, D. N., Lamb, P., Saul, M., and Winter-Nelson, A. E. (1997). Meanings of environmental terms. *Journal of Environmental Quality*, 26(3):581–589.

Jones, D., Molitor, D., and Reif, J. (2018). What do workplace wellness programs do? evidence from the illinois workplace wellness study. Working Paper 24229, National Bureau of Economic Research.

Jones, R. E. and Dunlap, R. E. (1992). The social bases of environmental concern: Have they changed over time? *Rural Sociology*, 57(1):28–47.

Lee, D. R. and Murphy, R. H. (2017). An expressive voting model of anger, hatred, harm and shame. *Public Choice*, 173(3-4):307–323.

Lehrer, S. and Xie, T. (2016). Box office buzz: Does social media data steal the show from model uncertainty when forecasting for hollywood? Working Paper 22959, National Bureau of Economic Research.

Lehrer, S. F. and Xie, T. (2018). The bigger picture: Combining econometrics with analytics improve forecasts of movie success. Working Paper 24755, National Bureau of Economic Research.

Lomasky, L. and Brennan, G. (1993). Democracy and decision: The pure theory of electoral preference. *Cambridge: Cambridge UP*.

Maniloff, P. and Mastromonaco, R. (2017). The local employment impacts of fracking: A national study. *Resource and Energy Economics*, 49:62–85.

Maur, A. T. (2015). Let's not frack this up: State-based solutions for the regulation of hydraulic fracturing and the disposal of flowback water. *Suffolk University Law Review*, 48(1):151–170.

Miao, H., Ramchander, S., Wang, T., and Yang, D. (2017). Influential factors in crude oil price forecasting. *Energy Economics*, 68:77–88.

Mueller, D. and Günther, L. (2003). C.(2003). *Public Choice III*.

Neto, A. B. F., Hodges, C. D., Pyun, H., et al. (2016). Voting dynamics and the birth of state-owned casinos in kansas. *Economics Bulletin*, 36(1):329–336.

Nowak, A. and Smith, P. (2017). Textual Analysis in Real Estate. *Journal of Applied Econometrics*, 32(4):896–918.

O'Roark, J. B. and Wood, W. C. (2011). Determinants of congressional minimum wage support: The role of economic education. *Public Choice*, 147(1):209–225.

Osborn, S. G., Vengosh, A., Warner, N. R., and Jackson, R. B. (2011). Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *proceedings of the National Academy of Sciences*, 108(20):8172–8176.

Peng, L., Meyerhoefer, C., and Chou, S.-Y. (2018). The health implications of unconventional natural gas development in pennsylvania. *Health economics*, 27(6):956–983.

Podolny, K. (2013). Gas transmission facilities: The limits on home rule. *Albany Law Review*, 77:705–1643.

Rabe, B. G. and Borick, C. (2013). Conventional politics for unconventional drilling? Lessons from Pennsylvania's early move into fracking policy development. *Review of Policy Research*, 30(3):321–340.

Rahm, D. (2011). Regulating hydraulic fracturing in shale gas plays: The case of Texas. *Energy Policy*, 39(5):2974–2981.

Raihanian Mashhadi, A. and Behdad, S. (2018). Discriminant effects of consumer electronics use-phase attributes on household energy prediction. *Energy Policy*, 118(November 2017):346–355.

Ritchie, A. (2014). On local fracking bans: policy and preemption in new mexico. *Natural Resources Journal*, 54(2):255–317.

Sardaro, R., Bozzo, F., and Fucilli, V. (2018). High-voltage overhead transmission lines and farmland value: Evidences from the real estate market in Apulia, southern Italy. *Energy Policy*, 119(April):449–457.

Scarcioffolo, A. R. and Etienne, X. L. (2018). How connected are the us regional natural gas markets in the post-deregulation era? evidence from time-varying connectedness analysis. *Journal of Commodity Markets*.

Silva, T. L., Morales-Torres, S., Castro-Silva, S., Figueiredo, J. L., and Silva, A. M. (2017). An overview on exploration and environmental impact of unconventional gas sources and treatment options for produced water. *Journal of environmental management*, 200:511–529.

Sovacool, B. K. (2014). Cornucopia or curse? reviewing the costs and benefits of shale gas hydraulic fracturing (fracking). *Renewable and Sustainable Energy Reviews*, 37:249–264.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

US Energy Information Administration (2017). *Annual Energy Outlook 2016 with Projections to 2040*. US Energy Information Administration, Washington DC.

US Energy Information Administration (2018). *Vermont State Profile and Energy Estimates: Profile Analysis*. US Energy Information Administration, Washington, DC.

Vidic, R. D., Brantley, S. L., Vandenbossche, J. M., Yoxtheimer, D., and Abad, J. D. (2013). Impact of shale gas development on regional water quality. *Science*, 340(6134):1235009.

Wegener, M. A. (2014). Drilling down: New York, hydraulic fracturing, and the dormant commerce clause. *BYU Journal of Public Law*, 28(2):351–372.

World Health Organization (2006). *Constitution of the World Health Organization*. World Health Organization, Geneva.

Yandle, B., Vijayaraghavan, M., and Bhattarai, M. (2002). The environmental kuznets curve. *A Primer, PERC Research Study*, pages 02–01.

Ziel, F. and Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, 70:396–420.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic-net. *Journal of the Royal Statistical Society*, 67(2):301–320.

Table 1: Data summary and source

Variables	mean	sd	min	max	data source
BIRTHPLACE	0.31	0.46	0	1	Votesmart
%COLLEGE	0.36	0.08	0.16	0.49	USDA
DEMOCRATIC	0.67	0.47	0	1	Followthemoney
%DEPENDENCY	0.55	0.06	0.46	0.65	ACS
DONATION($)	4310.74	6407.11	-2131	38894	Followthemoney
%DONATION_ENNR_PCT	0.02	0.07	0	0.64	Followthemoney
DONATIONPER ($)	104.8	369.56	0	3450	Followthemoney
ECONOMICS	0.02	0.13	0	1	Votesmart
ENVIRONMENT	0.04	0.2	0	1	Votesmart
HHINC	10.79	0.12	10.46	10.93	SAIPE
%HIGHSCHOOL	0.3	0.06	0.21	0.43	USDA
INCUMBENT	0.79	0.41	0	1	Followthemoney
MAGE	41.52	3.43	36.2	47.4	ACS
MP100F	97.06	1.86	93.7	102.2	ACS
MP100F18	94.97	2.43	90.2	101.9	ACS
%NOSCHOOL	0.08	0.02	0.06	0.16	USDA
POP	71515.04	48865.12	6306	156545	ACS
%POPA65	0.15	0.02	0.11	0.19	ACS
%POPU18	0.21	0.01	0.19	0.25	ACS
%POPU1824	0.1	0.03	0.06	0.15	ACS
%POPU2544	0.24	0.02	0.21	0.26	ACS
%POPU4564	0.31	0.02	0.27	0.37	ACS
%POVERTY	0.12	0.02	0.08	0.17	SAIPE
%POVU18	0.16	0.03	0.12	0.27	SAIPE
%POVU517	0.14	0.03	0.1	0.22	SAIPE
%SOMECOLLEGE	0.26	0.01	0.25	0.29	SAIPE
SUBJECT	0.06	0.24	0	1	Votesmart
%UNRATE	0.06	0.01	0.05	0.1	BLS
%WORKFORCE	0.65	0.03	0.61	0.69	ACS

N = 167.

Table 2: Model selection

Variables	Model Count	V8 268	V7 149	V12 124	V11 90	V24 65	V6 41	V16A 38
(INTERCEPT)	1000	1	1	1	1	1	1	1
DONATION ($)	1000	1	1	1	1	1	1	1
BIRTHPLACE	1000	1	1	1	1	1	1	1
DEMOCRATIC	1000	1	1	1	1	1	1	1
%POPU2544	969	1	1	1	1	1	1	1
ECONOMICS	928	1	1	1	1	1		1
%POVERTY	900	1	1	1	1	1	1	1
SUBJECT	779	1		1	1	1		1
MP100F	511			1	1	1		1
%DONATION_ENNR_PCT	479			1	1	1		1
HHINC	479			1	1	1		1
%SOMECOLLEGE	380			1		1		1
INCUMBENT	265					1		1
%POVU18	242					1		1
MP100F18	173					1		1
%DEPENDENCY	169					1		
%HIGHSCHOOL	157					1		1
DONATIONPER ($)	152					1		
%NOSCHOOL	146					1		
%UNRATE	117					1		
POP	114					1		
MAGE	111					1		
%POVU517	80					1		
%WORKFORCE	77					1		
COLLEGE	62							
%POPU4564	39							
%POPU18	4							
ENVIRONMENT	0							
%POPU1824	0							
%POPA65	0							

Table 3: Results

Variables	V8	V7	V12	V11	V6
	(1)	(2)	(3)	(4)	(5)
INTERCEPT	14.26***	14.45***	-221.33**	-176.87	15.15***
	(5.54)	(5.56)	(118.35)	(108.79)	(5.54)
	[1.27]	[1.28]	[-18.70]	[-15.00]	[1.36]
DONATION	0.00011**	0.00011**	0.00011**	0.00011**	0.00011**
	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)
	[0.000010]	[0.000010]	[0.000010]	[0.000010]	[0.000010]
BIRTHPLACE	-0.73	-0.75	-1.06**	-0.91	-0.82
	(0.53)	(0.53)	(0.58)	(0.56)	(0.52)
	[-0.06]	[-0.07]	[-0.09]	[-0.08]	[-0.07]
DEMOCRATIC	3.21**	3.25***	3.41***	3.28***	3.22***
	(0.56)	(0.56)	(0.62)	(0.59)	(0.55)
	[0.29]	[0.29]	[0.29]	[0.28]	[0.29]
%POPU2544	-35.46**	-35.96**	-95.93**	-82.97**	-37.11**
	(16.48)	(16.52)	(39.30)	(35.35)	(16.51)
	[-3.15]	[-3.19]	[-8.11]	[-7.04]	[-3.34]
ECONOMICS	2.02	14.95	1.88	1.88	
	(2414.74)	(1177.04)	(3880.42)	(3951.84)	
	[0.18]	[1.33]	[0.16]	[0.16]	
%POVERTY	-50.35***	-50.90***	-1.52***	-9.49	-53.57***
	(18.38)	(18.47)	(33.86)	(33.04)	(18.35)
	[-4.48]	[-4.52]	[-0.13]	[-0.81]	[-4.82]
SUBJECT	13.95		15.31	15.10	
	(1434.33)		(2178.67)	(2201.92)	
	[1.24]		[1.29]	[1.28]	
MP100F			0.38	0.41	
			(0.25)	(0.25)	
			[0.03]	[0.03]	
%DONATION_ENNR_PCT			2.32	2.77	
			(5.48)	(5.36)	
			[0.20]	[0.23]	
HHINC			18.46*	14.58*	
			(9.68)	(8.73)	
			[1.56]	[1.24]	
SOMECOLLEGE			29.47		
			(28.15)		
			[2.49]		
Accuracy	0.85	0.85	0.84	0.84	0.85
SD of Accurarcy	0.58	0.58	0.55	0.55	0.57
AIC	113.89	112.28	116.43	115.57	111.48
$R^2_{McFadden}$	0.46	0.46	0.49	0.48	0.45
$R^2_{Adj\ McFadden}$	0.44	0.44	0.46	0.46	0.44

***, **, * represents significant in 10%, 5% and 1% level of significance. Enclosed values in (.) are standard errors & enclosed in [.] are average marginal effects.