California Proposition 8: Voters Reject the "Fair Pricing" for the Dialysis Act☆

Tuyen Pham[1] **and Shishir Shakya**[2]

Abstract

In 2018, California held a ballot on proposition 8, the limits on dialysis clinics' revenue and require refund initiative. The proposition requires clinics to issue refunds to patients or their payers for revenue that exceeds 115% of the cost. In this paper, we explain the counties regional dynamics of the California proposition 8's failure. First, we develop a theoretical model on how voters weigh their cost and benefit to identify key determinants that might affect voting outcome. We then use voting data on Proposition 8 and employ simple OLS and LASSO approaches to estimate the effects of those key factors on real outcome.

☆ I am grateful to ____for helpful comments and guidance in this work.
[1] Tuyen Pham, College of Business and Economics, West Virginia University, Morgantown WV 26506 E-mail: tp0014@mix.wvu.edu
[2] Shishir Shakya, College of Business and Economics & Regional Research Institute (RRI), West Virginia University, Morgantown WV 26506 E-mail: shishir.shakya@mail.wvu.edu

1. Introduction

Chronic Kidney Disease (CKD) is any condition that causes reduction in kidney function over a period of time. End Stage Renal Disease (ESRD) is the last stage of Chronic Kidney Disease, where people lose about 85 to 90 percent of kidney function and have Glomerular Filtration Rate (GFR) of 15 or less. When CKD develops into ESRD, patients need dialysis to remove their body's liquid, salt, and wastes, what their kidneys now fail to do (National Kidney Foundation, 2015). Beside kidney transplant, dialysis become a lifesaving treatment for patients with ESRD. According to the National Kidney Foundation, an average life expectancy on dialysis is 5 to 30 years depend on their condition and how well they follow the treatment. There are 2 types of dialysis, hemodialysis (HD) and peritoneal dialysis (PD). Depend on patients' conditions such as how much waste in their body and how big they are, the time needed for each treatment is different. however, on average, a patient needs a treatment that lasts 4 hours and is done 3 times per week (National Kidney Foundation, 2015).

ESRD patients need to have dialysis their whole life and the cost of the treatment is expensive. According to United States Renal Data System (USRDS) annual report in 2018, an average cost for hemodialysis is $90,971 per person per year and an average cost for peritoneal dialysis is $76,177 per person per year. As of 2018, there are over 700,000 American patients are being treated for kidney failure including those who received kidney transplantation (Sullivan and Stern, 2018). Medicare covers most of the cost for dialysis. In 2018, Medicare spent $35.4 billion on fee-for service expenditure on for patients with ESRD, which accounts for over 7% of all Medicare expenditure (USRDS 2018 annual report).

In California, an average estimated cost per dialysis treatment in California is $250 per person with Medicare and $1000 per person with commercial insurance providers (Firozi, 2018). There are currently 663 dialysis clinic facilities in the state[3] that provide dialysis treatments for its' 80,000 patients every month (Firozi, 2018). Of those clinics, 316 dialysis clinics are operated by DaVita Clinics, Inc. and 132 clinics are operated by Fresenius Medical Care North America (FMC)[4]. Every

[3] and [4]: According to Medicare's list of Dialysis Facilities in the U.S. in 2018

year, those private clinics make billions of dollars in pretax profit, which a margin of 18% to 19% (Hiltzik, 2018).

On November 2018, the State of California placed a ballot on Proposition 8, Limits on Dialysis Clinics' Revenue and Required Refunds Initiative. The measure, which was sponsored by the Service Employees International Union – United Healthcare Worker (SEIU-UHW), would cap dialysis clinic facilities' profits at 15%. A "Yes" on Proposition 8 means dialysis clinics would be required to issue refunds to patients or patients' payers for revenue above 115% of the cost of direct patient care and healthcare improvement. A "No" vote means dialysis clinics would not have their revenue limited and would not be required to pay rebates.

Supporters argue that Proposition 8 would improve healthcare by forcing private dialysis to put higher percentage of their revenues into patient care. They also say that by putting a cap on clinics' revenue, it would reduce excess charges to insurance companies, thus lowering the premiums for all Californians. Opponents, including dialysis providers and their employees, argue that Proposition 8 will put dialysis at risk of being closed and thus reducing patient access to the treatment and putting lives of dialysis patients at risk.

While there was $19 million contributed to the supporting committees during the campaign, more than $111 million was contributed to the opposition committees, mainly through big dialysis providers[5]. The two dialysis giants, DaVita Inc. and FMC, who control the combination of nearly 67 percent of all dialysis clinic facilities currently operating in California, contributed more than $101 million to the opposition committees. Of that $101 million, more than 67 million dollars are from DaVita Inc., and the other $34 million are from FMC. During the campaign, they put out ads that say if the proposition is passed by voters, clinics will be at risk of being closed.

Despite the idea of the proposition is to protect dialysis patients and payers from being overcharged by private dialysis clinics, the measure was defeated. According to California's Secretary of State, 59.93% of voters voted to reject proposition 8. In this paper, we explain factors that impact voting outcomes.

[5] According to Ballotpedia.org, California Proposition 8

2. Theory model

If proposition 8 is enacted, there would be two main benefits. The first potential benefit, as mentioned above, if clinics want to keep bigger portion of their revenue, they must invest more money into direct patient care such as, which means better sanitation, better hygiene, and lower infection rate. The second potential avoid excessive charge from private clinics. Since the cost of dialysis treatments are mainly paid by Medicare, Medi-Cal, and private insurance companies, we argue that the first benefit is more of a direct benefit to individual voters. The lower costs would benefit payers, which are not patients in most of the cases. The biggest cost of passing proposition is the risk of reducing patient access to their lifesaving treatments due to clinics' closures.

When voter choose to vote between 2 alternatives, they weigh their own cost and benefit. If the benefit of passing proposition 8 is bigger than the cost, voters will be more likely to vote "Yes", otherwise, they will vote "No". Hence, voters weigh between improved healthcare and the cost of losing their access to the treatment.

$$B_i = prob_{having\_ESRD} \times b$$

$$C_i = (prob_{reducing\_access} \times c \times prob_{having\_ESRD})$$

$$U_i = B_i - C_i = U_i(prob_{having\_ESRD}, prob_{reducing\_access})$$

Assume the cost and benefit for all ESRD patients are the same.

Let $\boldsymbol{b}$ is each ESRD patient's direct benefit if proposition 8 is passed, which is the improvement in patient care. Voter $\boldsymbol{i}$, who can either have ESRD or not, so their benefit of having proposition 8 passed is $\boldsymbol{B_i}$. The benefit is determined by how likely an individual's probability to have ESRD. If voter live in an area where the incidence of ESRN is high, they would think their likelihood of getting the disease is high. When the incidence is higher, it also increases the benefit of voting "Yes" on proposition 8.

Let $\boldsymbol{c}$ is each ESRD patient's direct cost if proposition 8 is passed, which is the reduction in access to cares. Voter $\boldsymbol{i}$'s cost of having proposition 8 passed is $\boldsymbol{C_i}$. The cost for voter increases with the incidence of having ESRN and the probability of having clinic closures.

Let U_i is voter i's utility, which is the benefit minus cost. Since b and c are fixed, U_i is a function of the probability of having ESRN and the probability of having clinic closures. The factors that impact these probabilities are factors that impact voter's utility, hence, influence voting outcome. We expect to see the probability of having clinic closure having a negative effect of "Yes" votes. On the other hand, since the probability of having ESRN increases both voters' cost and benefit, we do not know if the net effect is negative or positive. However, since dialysis is a life-saving treatment, losing access to care has a serious harm on patients' health. We argue that the cost c is much bigger than the benefit b, which is the improvement in care. Therefore, the net effect is expected to be negative. This hypothesis can be test in out empirical approach.

3. Data

We use three different data sets for this project. The first data set is the voting outcome data from the California's Secretary of State. The data reports the numbers of voters, the numbers of Yes votes, the numbers of No votes.

The second data set we use is Dialysis Facilities in the U.S. This data set provides a list of all dialysis clinic facilities in California that registered with Medicare includes facilities' names, addresses, chain organization, and services provided. From this data, we calculate the number of clinics in each county, whether there are FMC clinics or DaVita's clinics in that county, as well as the numbers of FMC clinics and DaVita's clinics.

The third data set we use is California's 2018 County Health Ranking report. This data provides counties' demographic and economic characteristics. The data also reports different health outcomes by counties.

4. Empirical approach and results

4.1 OLS approach and results

In our first empirical approach, we chose to use simple OLS to test the theory model. Our dependent variable is the percentage of Yes votes in each county from California's Secretary of State. Our independent variable, as mention above, are factors that affect the county's kidney diseases incidence and the probability of having CKD clinic closure. Since DaVita Clinics' Inc.

and FMC spent millions of dollars on the opposition and said that they would close their clinics if they are no longer profitable. It might make voters think that the presence and the numbers of clinics from these two big chains in each county have big effects on the probability of having clinical closure. Hence, our first independent variables are the dummy variables that indicate whether there is any DaVita or FMC clinics in each county as well as the numbers of clinics from these two big chains. On the other hand, we include the numbers of dialysis clinics and dialysis station in each county in tour models because they also affect voters' voting behaviors. We do not have the county data of chronic kidney diseases' incidence. However, since people with high blood pressure and diabetes are more likely to develop kidney diseases than others, we use counties' percentage of obese and diabetes as other independent variables. We also control for other economic and demographic characteristics such as household income and the share of races in each county.

The primary results for this OLS approach are reported in table 1. Column 1 reports result from model 1, which includes the dummy variables indicate the existence of DaVita and FMC clinics, the total number of all clinics and dialysis station in each county. Model 2, reported in column 2, adds the numbers of DaVita and FMC clinics for each county. Column 3 shows results for model 3, which adds counties' percentages of obese and diabetes. Lastly, in column 4, we added control for the percentages of population with high school degree and/or college degree and the number of clinics per 100,000 population. All four models are controlled for demographic characteristics and economic characteristics.

We found that the presence of DaVita clinics in a county accounts for 5% less in "Yes" vote. The presence of FMC clinics in a county does not affect voting outcome. The finding is consistent with our theory model mentioned above. If there is DaVita clinics in a county, voters in that counties fear that they might lose the access to their care if DaVita closes their clinics when proposition 8 is passed. In model 3, we also found that one percent increase in the incidence of obesity decrease in each county decrease the percentage of "Yes" votes by 0.64 percent. Which confirms our hypothesis about ESRN incidence's negative net effect mentioned above in the theory model session. However, when controlling for education in model 4, this small negative effect become insignificant.

Table 1. OLS estimation

	Dependent variable:			
	% Yes			
	(1)	(2)	(3)	(4)
DaVita	-5.455**	-5.141**	-4.982**	-5.493**
	(2.175)	(2.351)	(2.304)	(2.691)
FMC	2.015	2.359	3.500	2.699
	(1.983)	(2.266)	(2.270)	(2.368)
# of facilities	0.402	0.511	0.553	0.721
	(0.679)	(0.738)	(0.717)	(0.751)
# of stations	-0.018	-0.018	-0.022	-0.028
	(0.030)	(0.032)	(0.031)	(0.032)
# of DaVita clinics		-0.188	-0.018	-0.055
		(0.433)	(0.432)	(0.430)
# of FMC clinics		-0.149	-0.389	-0.426
		(0.364)	(0.369)	(0.372)
High school rate				-0.039
				(0.066)
% College				0.125
				(0.187)
# facilities per 100,000				-0.815
				(1.078)
% Obesity			-0.637*	-0.506
			(0.366)	(0.379)
% Diabetes			-1.155	-0.895
			(0.986)	(1.037)
Household Income	0.0001	0.0001	-0.00002	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
% Non - English	1.503***	1.468**	1.378**	1.418**
	(0.536)	(0.558)	(0.542)	(0.564)
% Rural	-0.105**	-0.103*	-0.060	-0.119*
	(0.051)	(0.053)	(0.060)	(0.069)
Observations	58	58	58	56
R^2	0.583	0.586	0.630	0.657
Adjusted R^2	0.460	0.438	0.473	0.461
Residual Std. Error	5.149 (df = 44)	5.253 (df = 42)	5.084 (df = 40)	5.033 (df = 35)
F Statistic	4.732*** (df = 13; 44)	3.957*** (df = 15; 42)	4.013*** (df = 17; 40)	3.355*** (df = 20; 35)

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

4.2 Post LASSO approach and results

4.2.1 Post LASSO approach

In general, economic theory and intuition guide variable selections. In our case, general utilitarian theory and median voters' characteristics could define California Proposition 8 voting behavior. However, as a researcher, we do not observe exact data generating processes. Failure to adequately control can lead to endogeneity due to omitted variable bias. However, over-controlling leads to a loss of efficiency of estimates. A standard strategy is ad hoc to report the estimates implementing different sets of controls and show treatment effects are indifferent to changes in controls.

Causal interpretation relies on the belief that there are no higher-order terms of the control variables, no interaction terms, and no additional excluded variables that are associated both to treatment variable and outcome variable. Thus, controlling a large set of variables seems desirable to make this assumption more plausible. However, naively controlling the redundant variables reduces the ability to distinguish the impact of the interest variable and consequently produces less precise estimates. Moreover, including and controlling for all transformations of controls may not be feasible because the covariates space can increase as high dimensional, and regression is completely infeasible when the numbers of covariates exceed the number of observations in data. In our case, we have a "fat" dataset with 58 observations of 120 different variables. Under the assumption of sparsity, observables can be adequately controlled with the double-selection post lasso method proposed by (Belloni et al., 2013).

The double-selection post lasso method comprises three steps. First, run lasso[6] of the outcome variable on the list of potential control variables to select a set of predictors for the outcome variable. Second, run lasso of the variable of interest on the list of possible control variables to pick a set of predictors for the variable of interest. Third, run ordinary least squares regression of outcome variable on the variable of interest, and the union of the sets of regressors selected in the two lasso runs. Then correct the inference with usual heteroscedasticity robust ordinary least squares standard error — the estimates of the response of outcome variable on the interest variable

[6] The least absolute shrinkage and selection operator (LASSO) described in Tibshirani, (1996) is a machine learning algorithm for variable selection.

yield a causal interpretation. For the theoretical arguments see (Belloni et al., 2013, 2014; Nowak & Smith, 2017).

The primary goal is to inference about low-dimensional parameter from the high-dimensional nuisance parameter which comprises to solve auxiliary prediction problem quite well. Consider voting outcomes y_i as a partially linear model:

$$y_i = d_i\alpha_0 + g(z_i) + \xi_i, \quad E[\xi_i \mid z_i, d_i] = 0$$

$$d_i = m(z_i) + v_i, \quad E[v_i \mid z_i] = 0$$

where, we have a sample of $i = 1,\dots,n$ independent observation. d is policy/treatment variable possibly nonrandomly assigned an economic variable. In our study, it is the existence of DaVita. The α_0 is the target parameter of interest which answers the portion of voting variations due to the presence of DaVita. z_i is a high-dimensional vector of other controls or confounders. Based on the utilitarian voting framework and median voters' theorem, political, socio-economic and demographic features are in z_i. It is plausible to define that those features are a common cause for the existence of DaVita, and $m_0 \neq 0$, typically in the case of observational studies. We use linear combinations of control terms $x_i = P(z_i)$ to approximate $g(z_i)$ and $m(z_i)$. The list $x_i = P(z_i)$ could be composed of many transformations of elementary regressors z_i such as B-splines, dummies, polynomials, and various interactions. Having many controls poses a challenge of estimation and inference, therefore, to avoid such we assume the sparsity assumption that only a few among many variables in the z_i explains voting outcomes y_i.

$$y_i = d_i\alpha_0 + \underbrace{x_i'\beta_{g0} + r_{gi}}_{g(z_i)} + \xi_i$$

$$d_i = \underbrace{x_i'\beta_{m0} + r_{mi}}_{m(z_i)} + v_i$$

The sparsity then relates to $x_i'\beta_{g0}$ and $x_i'\beta_{m0}$ approximate $g(z_i)$, and $m(z_i)$ that requires only a small number of non-zero coefficients to render corresponding approximation errors r_{gi} and r_{mi}. An appealing method to estimate the sparse parameter from a high-dimensional linear model is the

Least Absolute Shrinkage and Selection Operator (LASSO) introduce by Frank & Friedman (1993) and Tibshirani (1996). LASSO simultaneously performs model selection and coefficient estimation by minimizing the sum of squared residuals plus a penalty term. The penalty term penalizes the size of the model through the sum of absolute values of coefficients.

Let's define a feasible variable selection via Lasso for outcome variable and policy/treatment variable. Here, we change the notation as the outcome, and policy/treatment variable take the following form

$$\tilde{y}_i = \underbrace{x_i\beta_1 + r_i + \varepsilon_i}_{f(\tilde{z}_i)}$$

$$\tilde{d}_i = \underbrace{x_i\beta_2 + m_i + \varepsilon_i}_{f(\tilde{z}_i)}$$

moreover, Lasso estimator is defined as the solution to:

$$\min_{\beta_1 \in \mathbb{R}^p} E_n\left[\left(\tilde{y}_i - \tilde{x}_i\beta_1\right)^2\right] + \frac{\lambda}{n}\|\beta_1\|_1$$

$$\min_{\beta_2 \in \mathbb{R}^p} E_n\left[\left(\tilde{d}_i - \tilde{x}_i\beta_2\right)^2\right] + \frac{\lambda}{n}\|\beta_2\|_1$$

where, the penalty level λ is a tuning parameter to regularize/controls the degree of penalization and to guard against overfitting. We choose λ by cross-validation in prediction. The $\|\beta\|_1 = \sum_{j=1}^{p}\left|\beta_j\right|$. The kinked nature of penalty function induces $\hat{\beta}$ to have many zeros, thus Lasso solution feasible model selection method. The estimated coefficients are biased towards 0; therefore, Belloni et al. (2011) suggest to run an OLS on selected variables also known as post-Lasso or Gauss-Lasso estimator.

Let $\hat{I}_1 = S\left(\hat{\beta}_1\right)$ denote support or the controls selected by feasible Lasso estimator $\hat{\beta}_1$ and $\hat{I}_2 = S\left(\hat{\beta}_2\right)$ denote support or the controls selected by feasible Lasso estimator $\hat{\beta}_2$. The post-double-selection estimator $\breve{\alpha}$ of α_0 is defined as the least squares estimator obtained by regressing y_i on d_i and the selected control terms x_{ij} with $j \in \hat{I} \supseteq \hat{I}_1 \cup \hat{I}_2$:

$$\left(\breve{\alpha},\breve{\beta}\right)=\min_{\alpha\in\mathbb{R},\beta\in\mathbb{R}^p} E_n\left[\left(y_i-d_i\alpha-\tilde{x}_i\beta\right)^2\right] \;:\; \beta_j=0,\forall j\notin\hat{I}$$

Belloni et al. (2013) provide theoretical results that the estimates are unbiased and consistent as:

$$\left(\left[\tilde{E}\tilde{v}_i^2\right]^{-1} E\left[\tilde{v}_i^2\tilde{\xi}_i^2\right]^{-1}\left[\tilde{E}\tilde{v}_i^2\right]^{-1}\right)^{-1/2}\sqrt{n}\left(\breve{\alpha}-\alpha_0\right)\xrightarrow{d} N(0,1)$$

4.2.2 Results from post-LASSO approach

Table 2 exhibits the primary results with five different model specifications. The dependent variable is the county level percentage of the population who voted in favor of California Proposition 8. Each model reports standard errors with HAC correction. Column (1) and (2) presents generic Lasso selection. Column (3) and (4) shows estimates from Lasso but always restricting on DaVita's presence. The column (5) and (6) shows estimates implementing double-selection post-Lasso, where, DaVita's presence is the independent variable of interest. Figure 1 exhibits, robustness for estimates or effect of DaVita's presence on the California Proposition 8 voting behavior.

Column (1) selects variables implementing generic Lasso from 19 variables dictionary. The estimates explain counties with higher population shares of Asian and native-born Americans are more likely to vote in favor while counties with higher under 18 years old population and diabetes population are less likely to vote in favor of California Proposition 8.

Column (2) selects variables implementing generic Lasso from 209 variables dictionary. This dictionary variable comprises 19 contemporaneous variables, their squared polynomials (19 more variables) and all the first level interaction (19*18/2 = 171). Column (2) estimates, counties with a non-working population (V4V8) and African American residing in rural are less likely to support while native-born American are more likely to support California Proposition 8.

The estimates in column (1) and (2) are explanatory only and cannot test channel of what precisely induces the voting behavior. In this research, we hypothesise that possibly the DaVita's presence in the county explain voting dynamics. One obvious choice is to model by selecting variable while always restricting the variable of interest. In this paper, our variable of interest in DaVita's presence in the county and column (3) and (4). However, in the robustness checks, we make an argument that this obvious looking procedure induces biases in the estimate.

Table 2: Double Selection Post LASSO, DaVita as Variable of Interest

	Percentage of voting in favor					
	(1)	(2)	(3)	(4)	(5)	(6)
DaVita			−2.553	−6.125***	−5.614**	−5.700**
			(1.916)	(1.873)	(2.228)	(2.544)
Household income	2.620		0.337			
	(4.761)		(4.254)			
% Under 18	−0.443*		−1.336***			
	(0.229)		(0.318)			
% Native	0.769***		1.009***			
	(0.238)		(0.225)			
% Asian	0.383***		0.188		0.334***	
	(0.117)		(0.120)		(0.118)	
% Non English			0.521**			
			(0.212)			
% Rural			−0.157***		−0.063	
			(0.046)		(0.040)	
% Diabetes	−1.359*		−0.089			
	(0.757)		(0.725)			
% Adult obese	−0.194	0.378	−0.055	0.244	−0.654***	−0.191
	(0.337)	(0.447)	(0.296)	(0.354)	(0.226)	(0.472)
$(\% \text{ Native})^2$		0.025**		0.032***		
		(0.012)		(0.010)		
Constant	29.295	55.216***	63.023	46.509***	55.659***	54.146***
	(57.506)	(8.119)	(51.594)	(5.039)	(6.228)	(7.083)
Observations	58	58	58	58	58	58
Variables	L1	L2	L1	L2	L1	L2
R^2	0.496	0.620	0.645	0.599	0.410	0.492
Adjusted R^2	0.437	0.529	0.578	0.534	0.353	0.357
F Statistic	8.365***	6.812***	9.676***	9.164***	7.214***	3.634***

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

Column (3) and column (4) show that the DaVita's presence in the county explains about 3.76 and 6.73 percentage of the vote in against California Proposition 8 respectively. Column (3) and (4) select variable from 19 and 209 variable selection dictionaries.

Column (5) and (6) selects variables based on the double-selection post lasso as described in the method section from 19 and 209 variable selection dictionaries. The variables selections are identical to each other. DaVita's presence in the county explains about 5.64 percentage less "Yes" votes on California Proposition 8.

4.2.3Robustness Checks

This research suffers power due to small observation. It seems the obvious choice to model by selecting variable while always restricting variable of interest as in column (3) and (4) estimates. However, unlike such, we show that our estimate in column (5) and (6) are less likely to behave erratically. We take 80 percent of the random selection of the full sample data and bootstrap variables selection for 10000 times then record the estimates of the effect of DaVita's presence. We then estimate the selected model with a full sample for 10000 times and record the estimates of the impact of DaVita's presence. We perform this procedure for the model presented in column (3) and (5) shown in Table 1. The Figure concludes the bootstrapping results.

Figure 1: Robustness checks of the effect of the DaVita's6 presence

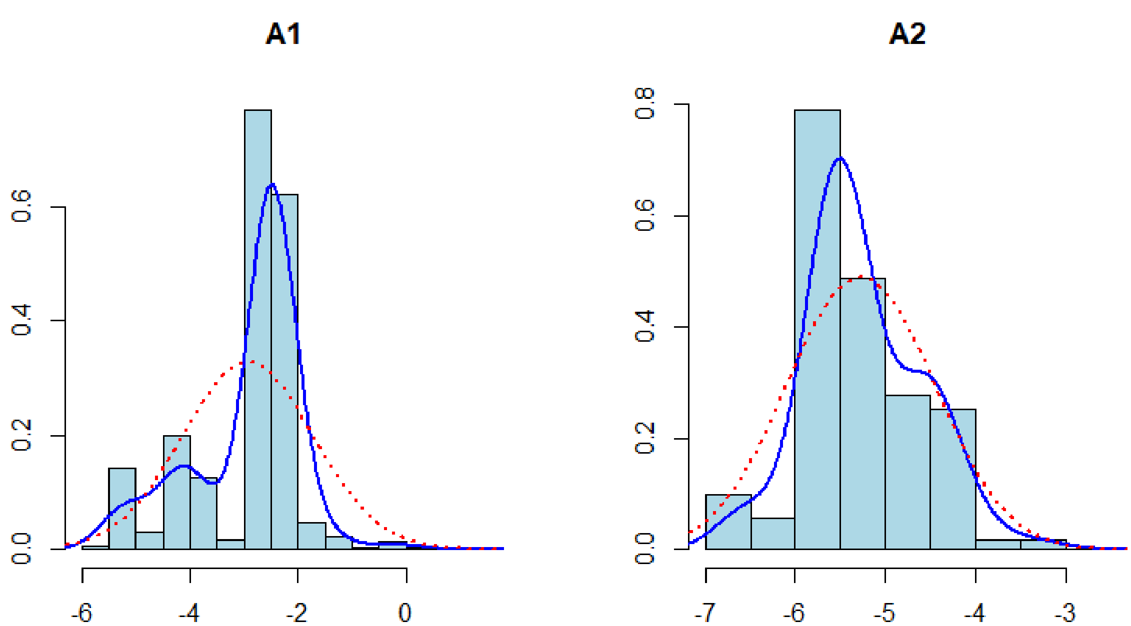
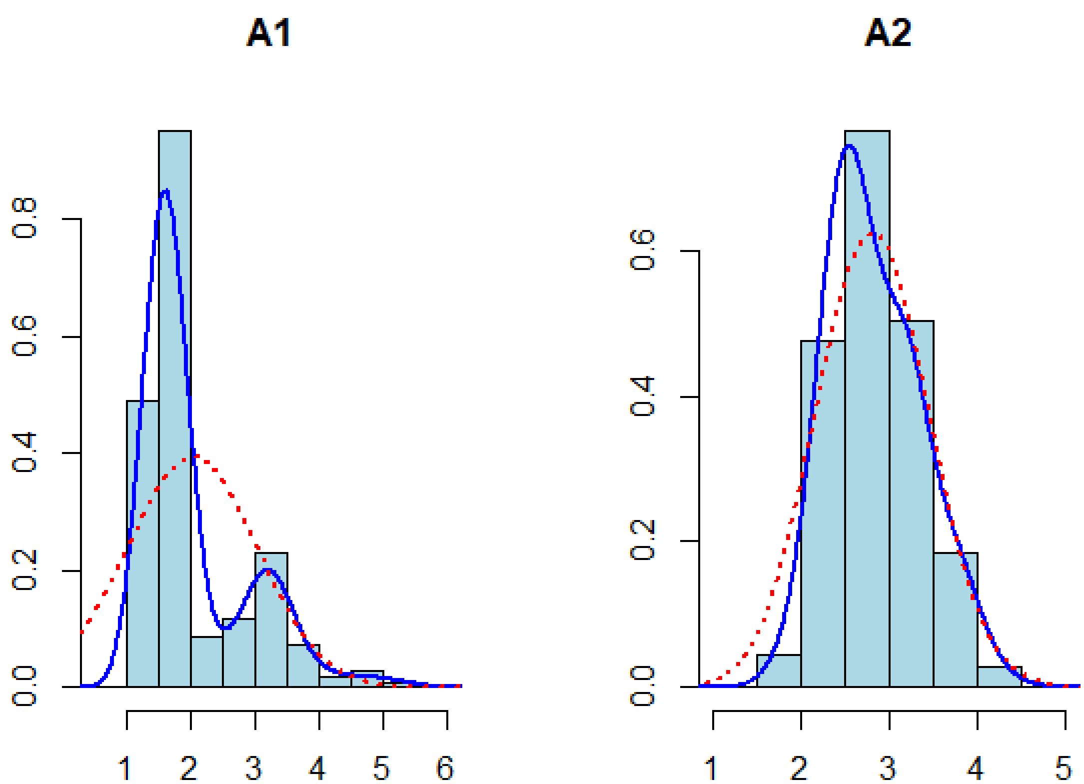


Figure 2: Robustness checks of the effect of the FMC's presence



Panel A1 and A2 show the density of impact of the DaVita's presence on voting behavior for the model presented in column (3) and (5) respectively. The density of effect of the DaVita's existence (in solid blue) is plotted against the normal distribution with the same parameters (in dotted red) respectively. Restricting DaVita's presence in Lasso selection yields relatively biased estimates (panel A1), while the double-selection post-Lasso produces relatively less biased and more consistent estimates (panel A2). Comparison of these plots suggests, that our main estimates presented in column (5) is likely to qualify a causal effect but do suffer low power.

5. Conclusion

The results from our 2 methods are consistent, it shows that the presence of DaVita clinics in a county accounts for 5% less in “Yes” votes on proposition 8. It also confirms the theory model mentioned above, the presence of DaVita clinics raises the voters’ risks of losing access to dialysis treatment. Since proposition 8 was meant to reduce the price of dialysis treatments, which are paid by Medicare, and the benefit of having better treatments is much less than the cost of losing access to cares, voters in counties with DaVita clinics tend to vote to reject the caps on dialysis clinics' revenues.

References

Belloni, A., Chernozhukov, V., & Hansen, C. (2011). Inference on Treatment Effects after Selection amongst High-Dimensional Controls. *ArXiv Preprint ArXiv:1201.0224*. https://doi.org/10.2139/ssrn.2051129

Belloni, A., Chernozhukov, V., & Hansen, C. (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, *81*(2), 608–650. https://doi.org/10.1093/restud/rdt044

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, *28*(2), 29–50. https://doi.org/10.1257/jep.28.2.29

End Stage Renal Disease in the United States. National Kidney foundation. https://www.kidney.org/news/newsroom/factsheets/End-Stage-Renal-Disease-in-the-US

Firozi, P. (2018). The Health 202: The dialysis industry spent more than $100 million to beat a California ballot measure.

Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*. https://doi.org/10.1080/00401706.1993.10485033

Hiltzik, M. (2018). Dialysis firms' profits are obscene. What will happen if California tries to cap them? Retrieved February 19, 2019, from https://www.latimes.com/business/hiltzik/la-fi-hiltzik-dialysis-20180720-story.html

Kidney Disease Statistics for the United States (2012). National Kidney and Urologic Diseases Information Clearinghouse. NIH Publication No. 12-3895, June 2012.

National Kidney Foundation (2015). https://www.kidney.org/atoz/content/dialysisinfo

National Kidney Foundation. (2017). Dialysis | National Kidney Foundation. Retrieved February 10, 2019, from https://www.kidney.org/atoz/content/dialysisinfo

Nowak, A., & Smith, P. (2017). Textual Analysis in Real Estate. *Journal of Applied Econometrics*, *32*(4), 896–918. https://doi.org/10.1002/jae.2550

Sullivan, J. D., & Stern, L. (2018). The State of Kidney Failure in the United States in 2018. *Urology and Nephrology*, *5*(1), 1–2. https://doi.org/10.19080/JOJUN.2018.05.555651

Tibshirani, R. (1996). Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B*. https://doi.org/10.2307/2346178

USRDS. (2018). 2018 Annual Data Report. Retrieved February 10, 2019, from https://www.usrds.org/adr.aspx

2018 Annual Data Report Highlights. United States Renal Data System. https://www.usrds.org/adrhighlights.aspx