Essay - Data Analytics - Topics 1 : The Estimation of the Relation between Agricultural output and Electoral Vote

Davi Méaille

August 2022

Contents

A - Introduction

Marx used to say that farmers lack of a class consciousness in France (Marx (1852)). For that reason, they were not the ideal ground where to start a potential revolution, as they tend more to vote for conservative parties. Over the twentieth century, we could even assume that some of them have become capitalists, as their properties and the benefits they derived from them have gotten larger. However, this does not necessarily mean that they are not interested by social questions at all, especially when confronted to it. For that reason, we study here the relation between farms performance and electoral behavior in the USA.

We will focus mainly on the opposition between republican and democrats, as they can represent an easy opposition between more or less interest for social issues, even though it is not completely adequate. We will use votes both for congressional and for presidential elections, as both types of election can represent different stakes. We will use a proxy for farm performance, average value per acre of farmland in a county, that we will relate to votes.

One issue that may arise is the one of identification. Running a simple panel within estimator may highlight more a correlation than a real impact. Indeed, actual vote may interfere with actual farm performance, but the contrary is also true, as economic conditions constitute a determinant of votes (Erikson (1989)), even though not necessarily the only one (Bourdieu (2016)). To account for this, we will run our model without lag and with a lag of four years, assuming that voting can have caused a change four years later. It will also be consistent with our hypothesis that the opinions in years of election are the same the following years without election.

B - Data

B -.1 Sources

We use agricultural and electoral data from the Inter-university Consortium for Political and Social Research (ICPSR). We use United States Agriculture Data (ICPSR 35206) and Electoral Data for Counties in the United States: Presidential and Congressional Races, 1840-1972 (ICPSR 08611).

B -.2 Description

B -.2.1 Agricultural Data Description

Once we have prepared our agricultural data, the data set that we obtain is summarized in Table 1

Table 1: Summary statistics for our agricultural data

Statistic	N	Mean	St. Dev.	Min	Max
FIPS	42,405	30,451.460	14,919.610	1,000.000	56,045.000
TOTPOP	42,379	81,817.500	990,568.900	0.000	179,323,175.000
AVERAGE AREA	40,636	436.995	3,266.416	0.000	450,000.000
AVERAGE VALUE PER ACRES	41,886	90.359	756.733	0.000	85,741.000
YEAR	42,406	1,929.798	24.548	1,880	1,964

where TOTPOP is total population in a given year and a given county and AVERAGE AREA is the average size of farms in a county and AVERAGE VALUE PER ACRE is average value per acre of farmland in a county. The latter variable is our proxy for farm performance.

B -.2.2 Electoral Data Description

Each two years, citizens have to vote for one-third of the seats in the American Senate and each four years, they have to vote for the House of Representatives. This gives us thus congressional electoral results for each two years and presidential results for each four years. Our data set is thus composed of years between 1840 and 1972, for a total amount of more than 200,000 observations, summed up in the following Table 2.

Table 2: Summary statistics for electoral data

Statistic	N	Mean	St. Dev.	Min	Max
FIPS	426,790	31,476.850	15,265.080	1,010	59,400
CONG DEM	426,790	278.045	403.460	0.000	999.990
YEAR	426,790	1,906.500	38.681	1,840	1,973
CONG REP	382,200	222.530	379.119	0.000	999.990
PRES DEM	426,790	244.124	383.654	0.000	999.900
PRES REP	375,830	189.485	345.070	0.000	999.900

CONG DEM represents the share of congressional vote obtained by democrats. CONG REP represents the share of congressional vote obtained by republicans. PRES DEM represents the share of presidential vote obtained by democrats. PRES REP represents the share of presidential vote obtained by republicans. We did not remove the meaningless data in our data set, because doing so could lead to the removal of meaningful data for other variables in the same counties. We did it,

however, in the regressions.

B -.3 Construction of variables

All our variables are computed on a county-basis.

We also need to present how the FIPS codes are computed. Each state in the US has its own postal and FIPS codes, that allow to identify it. Each county within each state has then its own FIPS code as well. The total FIPS code for a given county is built by writing the state FIPS code multiplied by 1000 and by adding the FIPS code of the county within the state.

B -.4 Final data set

Once merged over meaningful years that are present in our agricultural data set, we obtain a data set that is summarized by Table 3.

Table 3: Summary statistics for our final data set

Statistic	N	Mean	St. Dev.	Min	Max
TOTPOP	41,581	39,244.350	139,051.700	0.000	6,038,771.000
AVERAGE.AREA	39,872	438.889	3,296.247	0.000	450,000.000
AVERAGE.VALUE.PER.ACRES	41,087	90.749	763.960	0.000	85,741.000
CONG.DEM	41,606	105.684	211.532	0	1,000
CONG.REP	41,606	88.376	215.296	0	1,000
PRES.DEM	41,606	71.529	140.115	0.000	999.900
PRES.REP	41,606	65.267	140.950	0.000	999.900

B -.5 Data preparation

To match the years of election with the years at our disposal for farm performance, we assume that the opinions during the years without election are the same as the last year featuring an election. Since people are voting each two years for congressional election, we just use each vote at year t also for year $t+1$. For presidential election, we use vote at year t for the three following years.

We start at year 1880 and we include years up to until 1960 because the data sets at our disposal from ICPSR 35206 don't include all our control variables for years before and after that.

C - Estimation framework

I will use a regular individual fixed effect to account for county-fixed effect (as counties count as our individuals in our framework). We will thus estimate the following equation:

$$Y_{i,t} = \alpha_i + \beta \cdot VOTE_{i,t} + X_{i,t} + \varepsilon_{i,t} \tag{1}$$

where $Y_{i,t}$ is farm performance, α_i is county-fixed effects. $VOTE_{i,t}$ is our variable for vote. It can be either vote for democrats in congressional elections, for republicans in congressional elections and the same for presidential elections. Eventually, $X_{i,t}$ represents a matrix of additional controls variables, including average farm size in the county and total population of the county, as they can be decisively related to farm performance.

We estimate it using the within-estimator to capture within individual variations over time. We will perform the same regression a second time by adding a lag of two or more years to vote to see if we can capture a potential causality of vote on farm performance. This is what come closest to the estimation of a real effect of vote on farm performance. This framework will allow to pretend more to capture causality: it is unlikely that actual farm performance influences past votes, whereas it is possible that it influences actual votes.

D - Results

D -.1 No lag

We have thus performed the same regression for our four different vote variables. The results are presented in Table 4.

We can make several observations. First of all, the coefficients are only relevant for presidential elections. Second, We do have coefficients of contrary signs for republican and democrats elections, suggesting that our regressions have capture the opposition between both of them. This allows us to say that counties that vote more for republican candidates rather than democrat candidate tend to be counties where average value per farm land acres is higher. The trend is even more significant when it comes to the influence of voting for republicans than for democrats.

Table 4: Regressions Coefficients and Statistics

	Dependent variable:			
	AVERAGE.VALUE.PER.ACRES			
	(1)	(2)	(3)	(4)
CONG.DEM	0.413			
	(0.269)			
CONG.REP		0.316		
		(0.267)		
PRES.DEM			-0.376^{*}	
			(0.227)	
PRES.REP				0.615^{**}
				(0.265)
AVERAGE.AREA	-0.0001	0.00004	0.00001	-0.0002
	(0.003)	(0.003)	(0.002)	(0.002)
TOTPOP	0.002^{***}	0.002^{***}	0.001^{***}	0.001^{***}
	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Observations	33,061	37,376	38,872	38,931
R^2	0.049	0.048	0.034	0.038
Adjusted R^2	-0.031	-0.032	-0.046	-0.043
F Statistic	525.258^{***} (df = 3; 30478)	580.323^{***} (df = 3; 34478)	423.413^{***} (df = 3; 35874)	468.156^{***} (df = 3; 35928)
Note:				*p<0.1; **p<0.05; ***p<0.01

D -.2 With lags

The results are presented in Table 5 below. Again, we can make several observations. First, the significance of the coefficients has increased, meaning that we have capture a more precise estimation of our parameter. Second, the coefficients that are now significant are of higher magnitude than they were before.

The main point of using lag is to pretend to disclose a causality link, that is: the counties that have voted for a given party during some elections will see some result on their performance a few years later.

It is however surprising that average value per acre is correlated positively both with democrat and republican votes for congressional elections, even if the latter is not significant. We do not have an explanation for this result yet.

The R^2s remain however very low. The relations are thus significant, but the votes four years in the past only explain a very little variation of average value per acre.

E - Limits of our results

E -.1 Limits of our proxies

A first limit pertains to the choice of our proxy for farm performance. The choice of average value of farmland and buildings per acre in dollars of the year entails first the problem that it includes the trend of inflation. This is however captured by the within-estimator, that considers within individual variations. Another issue is pertains to what factors can influence this average value other than farm performance, making it a misleading proxy variable. Indeed, this value can be influenced by parameters that are captured by individual (that is county) fixed effects, like geography, type of soil (mountainous for instance) aso... However, other parameters like the differentiated impact of a crisis on all region are also to be taken into account: first, a recession does not affect all counties in the same way, but a petroleum crisis will also affect more counties further away from counties that have their own resources. Eventually,

E -.2 Causality

There is also the classic issue of causality versus correlation. In our case, it is all the more important than we have not been very specific concerning the channel of transmission, that is the mechanism, relating farm performance to vote. As a matter of fact, without lags, there could also be a mechanism

Table 5: Regressions Coefficients and Statistics with Lag of Four Years

	Dependent variable:			
	AVERAGE.VALUE.PER.ACRES			
	(1)	(2)	(3)	(4)
CONG.DEM	0.865***			
	(0.260)			
CONG.REP		0.284		
		(0.277)		
PRES.DEM			−0.547**	
			(0.240)	
PRES.REP				0.869***
				(0.281)
AVERAGE.AREA	−0.0003	−0.0001	0.0004	−0.00004
	(0.003)	(0.003)	(0.003)	(0.002)
TOTPOP	0.001***	0.001***	0.002***	0.001***
	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Observations	32,637	36,636	38,548	38,613
R^2	0.036	0.030	0.041	0.036
Adjusted R^2	−0.045	−0.051	−0.039	−0.044
F Statistic	376.117*** (df = 3; 30092)	352.506*** (df = 3; 33804)	511.392*** (df = 3; 35574)	447.570*** (df = 3; 35636)
Note:				*p<0.1; **p<0.05; ***p<0.01

relating vote to farm performance : for instance, it could be that farmers with higher farm performances are more interested in voting for a party that would protect their benefits with less social care. It could also be that farms with higher performance hire more workers and are thus more concerned with social laws that would facilitate the production by ensuring the welfare of the workers. We are thus not sure to isolate a univocal effect of votes on farm performance.

To measure this more precisely, it would be interesting to relate the vote to the implementation of some types of policies (for instance lower worker costs) and then relating this sort of event to farm performance. This would highlight a more concrete mechanism.

Eventually, there is no concrete mechanism in our study because we have given no plausible behavior of agents underlying the equations that we have tried to estimate. This is due to the fact that our equation does not identify any mechanism enough.

F - Conclusion

We wanted to determine if there was a significant relation between farm performance and vote in the USA. We have performed a simple individual fixed-effect regression, relating average value per acre of farmland per county to different types of electoral results. Our results suggest that the relation is significant when using variables for presidential election but not for congressional election. This suggest that the polarization for presidential election is more important than for congressional election, where a lot of local factors can be at stake. More importantly, we have found counties with higher average value per farmland acres to vote more for republican candidates at presidential election. This result points toward the fact, that they are more conservative and probably closer to capitalistic behaviors. Eventually, our main results show that the significance of the coefficients equation that we wanted to estimate is higher when we add lags, pointing toward an effective relation, even though very tight.

References

Bourdieu, P. (2016). *La distinction: critique sociale du jugement*. Minuit.

Erikson, R. S. (1989). Economic conditions and the presidential vote. *American Political Science Review*, 83(2):567–573.

Marx, K. (1852). The 18th brumaire of louis napoleon bonaparte.