Two Methods for Detecting Fraud In Mass Elections Using Official Election Returns

Misha Myagkov and Peter C. Ordeshook

University of Oregon and The California Institute of Technology

1. Introduction

In 2004, both Russia and Ukraine held presidential elections that provided an unprecedented opportunity to explore alternative methodologies for detecting and measuring election fraud in developing democracies. First, in Russia, Vladimir Putin's reelection was, with or without fraud, a foregone conclusion and not even the head of the Communist Party, Gennady Zyuganov, sought to compete against him. One might suppose, then, that such a circumstance would minimize the necessity for stuffed ballot boxes, manipulated vote counts and election protocalls whose numbers bear little relation to ballots cast. However, we need to keep in mind the internal dynamics of Russian politics and the incentives of regional elites to curry favor with authorities in Moscow. With the ultimate winner not in doubt, those elites were, in effect, in competition with each other for who could demonstrate greatest fealty to the powers that be in the Kremlin. In the 2000 election, in contrast, the ultimate outcome remained in doubt through much of the campaign. And in a state where historically backing the looser rarely led to a healthy outcome, those same regional elites would reasonably feel restrained in exerting an all-out effort for one candidate or another. But in 2004, such restraint no longer existed.

A second feature of Russian politics should also be kept in mind. There are, in fact, two 'Russias' – its oblasts and its ethnic republics. Whereas a healthy competition for control of regional governments often characterized regional politics thereby ensuring that regional

governors and the like held imperfect control of the levers of power within their domain, the politics of the ethnic republics such as Tatarstan, Dagestan, and Baskirostan resembled autocracies, with their presidents commonly winning reelection with numbers that echoed a Soviet past (i.e., winning vote shares of 95% percent and turnout in excess of 98%). Thus, if there were regional authorities equipped to exert full control over electoral processes and outcomes, those authorities were to be found in the ethnic republics. This constellation of factors leads to the reasonable hypothesis that

(*Hypothesis 1*) *whatever the degree of fraud can be said to characterize Russian elections, that fraud would be more pervasive in 2004 than in 2000 (although not necessarily nonexistent in 2000), and it would be most evident in the ethnic republics than in Russia's oblasts.*

Ukraine, in 2004, presented a different scenario. Here the main competitors were the West-leaning Vicktor Yushchenko and the Russian-backed Vicktor Yanukovich. It was also evident early in the campaign that despite his Russian backing (including at least one campaign appearance by Putin on his behalf) and the support of Ukraine's incumbent president and administration (and all the advantages of money and media access that such backing afforded him) Yanukovich would not surpass the 50 percent threshold required to avoid a runoff against his opponent. Thus, although some degree of fraud might have characterized the first round in October, all special efforts at 'mobilization' were reserved for the second and ostensibly decisive round in late November. There is little need at this point for us to recount events that followed that second round, except to recall the half million or so Yushchenko supporters who camped out

in Kyiv's Independence Square protesting a stolen election, the labeling of the outcome as illegitimate by innumerable heads of state (including political leaders in the Scandinavian states, Poland, Canada, the European Union and the United States) and finally the decision of Ukraine's Constitutional Court that there was sufficient evidence of fraud to invalidate the November outcome and a call for a rerun of that election in December. What followed was a relatively massive influx of journalists, international observers, and election monitors as well as the abandonment of Yanukovich by both his Russian fellow travelers and Ukraine's administrative hierarchy. In short, Ukraine offered researchers in election fraud an almost perfect social science experiment: Two elections, one month apart, with the same electorate and the same two candidates, but with considerably less opportunities and incentives on the part of regional authorities to engage in the fraudulent machinations that characterized the November contest. Thus, we can formulate a second hypothesis; namely,

(Hypothesis 2) whatever indicators of fraud we find in Ukraine's November presidential runoff should disappear or at least diminish in significance in the December re-run of that runoff.

There is one final hypothesis that can be gleaned from these events and circumstances. Specifically, many of the persons who directed or helped direct Putin's reelection campaign early in 2004 soon thereafter traveled to Kyiv to 'facilitate' Yanukovich's campaign. This, and the fact that although Russia and Ukraine are now independent states, they share a Soviet past and, thereby, the same tools for implementing fraud when fraud is on the agenda. This then yields a third hypothesis:

(Hypothesis 3) Whatever indicators of fraud can be found in official election returns in Russia in 2004, should make their appearance as well in Ukraine's November runoff election (and vise versa).

This essay begins with a brief summary of our earlier efforts at exploring these three hypotheses, but we do so as a preliminary to looking at data from the United States and from Mexico's recent controversial presidential contest. We want to consider whether any of the indicators of fraud found in Russia and Ukraine make their appearance in the United States – notably in the 2000 presidential election in Florida and the 2004 election in Ohio – and whether data from the United States offers any warnings about the general applicability of our methods as applied to Russia and Ukraine. And finally we want to consider whether those same indicators might compel us to explore specific sources of fraud, if any, in Mexico.

2. Russia and Ukraine, an Overview

Because we view our research as augmenting what on-the-ground election observers might conclude, our analysis relies exclusively on official election returns aggregated up to the level of individual polling stations (precincts) or rayons (counties). The core issue we address, then, is whether we can detect election irregularities in these official returns consistent with fraud and with what we know *a priori* about the elections considered. Of course, in some instances, fraud is self-evident. Consider the rayon in Tatarstan in 2004 that consisted of forty one polling stations, wherein no station reported turnout below 95 percent, where none gave Putin less that

98 percent of the vote, and where twenty four of those stations reported 100 percent turnout and 100 percent of the vote for Putin. Or a second rayon in Tatarstan in which all forty four stations reported 100 percent turnout, where all but one reported 100 percent of the vote for Putin, and where the one `defector' reported 100 percent of the vote for the candidate adjacent to Putin on the official rayon protocall (clerical error?).

Such examples illustrate that fraud is sometimes committed shamelessly in parts of Russia with little regard for Western opinion (or with the understanding that other political considerations will lead observers to anoint any election as free and fair). But such examples cannot be relied on in other states or used to estimate fraud's overall magnitude, even in Russia. Hence, in the case of Russia, and subsequently Ukraine, three alternative indicators are considered wherein before we deem an election suspect all three must simultaneously be consistent with our three hypotheses. These indicators are (1) estimates of the flow of votes from one election to the next that 'make sense', by which we mean that shares fall in the interval [0, 100%] and do not indicate that one candidate or another suddenly and inexplicably received an inordinate share of support from those who previously had been nonvoters; (2) distributions of turnout that are approximately normally distributed across precincts or counties, and that are otherwise not consistent with the hypothesis that some subset of precincts have had their turnout artificially augmented with stuffed ballots or falsified election totals; and (3) a logical relationship between turnout and a candidate's share of the eligible electorate, where by 'logical' we mean that if turnout increases, then *ceteris paribus*, a candidate (or party) should share in this increase or at least not suffer from it.

Insofar as this third indicator is concerned, suppose we estimate the regression

*Candidate i's share of the eligible electorate = A*Turnout + B*

Then in a normal election in otherwise homogeneous districts (where turnout varies as a function of factors that do not normally correlate with a candidate's level of support) the coefficient A should approximate that candidate's overall share of the vote and B should be close to zero. That is, if a candidate wins on average say 65 percent of the vote, then *ceteris paribus* for every additional 100 votes, he or she should receive 65 additional votes.

Because in this essay our research is directed, at least in part, at data from the United States, we forgo the first indicator since, when examining individual states, it is often impossible to conduct a flow-of-votes analysis owing to near continuous redefinition in the geography of precincts (see Myagkov, Ordeshook and Sobyanin 1997, Myagkov and Ordeshook 2001, 2005, and Myagkov, Ordeshook and Shakin 2005 for a flow of votes analysis in Russia and Ukraine). Focusing then, on our second and third indicators, let us begin with Figure 1, which graphs the distribution of turnout in both oblasts and republics for both the 2000 and 2004 presidential elections in Russia (from Myagkov, Ordeshook and Shakin 2005: Figure 4, p. 99). The first thing to note is the clear differentiation in the form of these two distributions. Oblasts in 2000 offer an almost a perfect example of a normal distribution whereas in 2004, we witness a marked drop in overall turnout (consistent with the election being a forgone conclusion) but with a tail that inexplicably suggests that some oblast rayons stubbornly maintained high rates of turnout. The data from Russia's republics, on the other hand, exhibit no ambiguity in interpretation. The data there for 2000 are wholly consistent with the hypothesis that although turnout absent fraud would be normally distributed across rayons, a significant subset of those rayons have had their turnout artificially augmented so as to generate a bimodal distribution. And in 2004, this

augmentation is of such magnitude as to shift the mode of the entire distribution to the right. Clearly such data is wholly consistent with Hypothesis 1.

[Figure 1 here]

Now let us consider our third indicator – the relationship between turnout and a candidate's share of the eligible electorate. First, let us consider by way of a simple example how fraud can distort an otherwise normal relationship between turnout and a candidate's share of the eligible electorate. Imagine a simple six precinct election between two candidate, A and B. Suppose 100 people live in each precinct and that A's support is 25% across the precincts while B's is 75%. Suppose also that in a regular case and for wholly exogenous reasons (e.g., weather) three of these six precinct would have a turnout of 40% (40 people vote), and the remaining three a turnout of 60% (60 people vote). Now imagine that in two of the three 40% turnout precincts that 40 ballots are fraudulently added to candidate B's total, thus raising turnout to 80%. If we were now to graph turnout against A and B's absolute vote, we have three clusters of points for each candidate:

- One (regular) data point per candidate at .4 turnout with A's share of the eligible electorate at .10 and B's at .3 (i.e., the points A: (.4,.1) and B: (.4,.3))
- Three (regular) points per candidate at .6 turnout with A's share of the eligible electorate at .15 and B's at .45, or equivalently, the data points for A (.6,.15), and for B (.6,.45)
- Two (fraudulent) points at .8 turnout; for A (.8,.1) and for B (.8,.7).

Since with fraud there is more weight (more points) on the highest value of observed turnout as compared to the lowest value (two points at .8 vs. one point left at .4) OLS will estimate a

relationship between turnout and absolute vote that exceeds 1.0 for B and is negative for A. Notice, moreover, that we can simplify the example if we assume that not only are votes added to B, but that some votes are stolen from A. Consider just two precincts, both with 40% turnout and suppose that in one of them 40 new votes are added to B's total while 5 are stolen from A's, which are also credited to B.. Then the high turnout precinct will have the following coordinates: For A, (.8,.05) and for, B (.8,.75). The low turnout (regular) precinct will correspond to the data points for A, (.4,.1) and for, B (.4,.3). If we now regress turnout against absolute vote share, the coefficient for B will now be (.75-.3)/(.8-.4)=.45/.4=1.25 whereas the coefficient for A will be (.05-.1)/(.8-.4) = -.125.

Turning now to the case of Russia, consider Tver and Samara oblasts, which are generally known to be a relatively competitive regions only weakly controlled by their regional governors (and where the governor of Tver, running as a candidate for United Russia -- the Kremlin's party -- recently lost his bid for reelection). In Tver in 2004 the estimated regressions for Putin and Haritonov were (for the actual graphs of these regressions see Myagkov, Ordeshook and Shakin 2005, Figures 6A-D, pp. 102-3):

$$\text{Putin}_{\text{Tver}} = 0.842\text{Turnout}_{\text{Tver}} - 0.075$$

$$\text{Haritonov}_{\text{Tver}} = 0.213\text{Turnout}_{\text{Tver}} - 0.028$$

and in Samara we estimate

$$\text{Putin}_{\text{Samara}} = 0.755\text{Turnout}_{\text{Samara}} - 0.074$$

$$\text{Haritonov}_{\text{Samara}} = 0.274\text{Turnout}_{\text{Samara}} - 0.042$$

Now, however, consider the two especially suspect republics of Tatarstan and Bashkortostan. Here we get

$$Putin_{Tatarstan} = 1.673Turnout_{Tatarstan} - 0.711$$

$$Haritonov_{Tatarstan} = -0.247Turnout_{Tatarstan} + 0.263$$

and

$$Putin_{Bashk} = 1.413Turnout_{Bashk} - 0.433$$

$$Haritonov_{Bashk} = -0.204Turnout_{Bashk} + 0.214$$

Thus, not only does Putin win in excess of 100 votes for every 100 voters arriving at the polls, but his communist party challenger actually loses votes!

For reasons we discuss later, such indicators, taken by themselves, ought to be treated with care. However, let us now briefly consider Ukraine. Figure 2 summarizes four distributions of turnout – turnout in both the second and third rounds of voting in 2004 after separating rayons into those carried by Yanukovich versus Yushchenko (from Myagkov et al, 2005, Figure 7B, p. 111). What we see here is a wholly unexceptional (normal) distribution in the second round for those rayons carried by Yushchenko whereas the distribution for Yanukovich in that round looks much like Russia's republics in 2000. In round three, however, that distribution returns to the normal form we expect from a fraud free contest, whereas now if there is an anomaly to report, it occurs within Yushchenko's support base. These data are consistent, then, with the hypothesis that the influx of observers into Yanukovich's base, Eastern Ukraine (along with his abandonment by the powers-that-be in Moscow and Kyiv), discernably suppressed fraud there, but the attention paid to Eastern Ukraine in this round allowed Yushchenko's supporters in Western Ukraine the opportunity to afford their candidate some added 'assistance'.

[Figure 2 here]

Our final sequence of figures relating to Ukraine concerns the relationship between turnout and a candidate's absolute vote, and here it is useful to contrast rayon level data from two specific regions – Lviv and Donetsk. Donetsk is Yanukovich's home region and the oblast he governed before entering the national political stage. Lviv, in contrast, is Ukraine's western-most city and the most Western in its outlook. Figures 3a and 3b plot each candidate's absolute vote against turnout in each round for each of Donetsk's 23 rayons and Lviv's 12 rayons (from Myagkov et al, 2005, Figures 13A, B, p. 124). The first thing to note is that neither candidate receives much of the vote in their 'alien' region. But consider how each candidate's absolute vote share changes in general as we move from round one, to two, to round three. In Lviv (Figure 3b) Yushchenko's vote increases only incrementally as we move from one round to the next; and gioven that Yanukovich's vote increases not at all as turnout increases (and in a statistically insignificant way, actually decreases), we might speculate that each round of voting saw some additional albeit slight less-than-legitimate efforts there on Yushchenko's behalf.

[Figure 3a and 3b here]

The data in Donetsk stand in sharp contrast to what we see in Lviv (Figure 3a). Here essentially all increase in turnout from round 1 to round 2... an increase that often exceeded 20 percent -- benefits Yanukovich. But as we move to round 3 – the round heavily monitored by election observers – our data points "slide back" as if beads on a string to nearly mirror the turnout-vote totals found in round 1. Once again, although this pattern does not 'prove' fraud, it is wholly consistent with a substantial stuffing of ballot boxes in round 2 to benefit Yanukovich alone, and the near absence of such fraud in round 3 (for a flow of votes analysis that sustains our hypothesis of significant fraud benefiting Yanukovich see Myagkov et al 2005, Table 8, p.

120).

3. California, North Carolina and Florida

Not every state and not every county in a state reports data that allows for easy access to precinct-level election returns. In this section, then, we consider a set of non-randomly chosen states and counties, beginning with the county of San Francisco and the 2004 Bush-Kerry presidential contest. What is interesting about the data here is that we are dealing with a county that voted overwhelmingly for Kerry in much the same way as Russia's republics voted for Putin in 2004 and the eastern regions of Ukraine voted for Yanukovich. However, looking at Figures 4a and 4b, notice first in Figure 4b the wholly normal (i.e., non-suspicious) distribution of turnout across precincts. Figure 4a, in turn, shows that despite President Bush's general unpopularity in that county, even he gained a few votes with every increase in turnout (5 for every 100 additional voters). Minimally, then, the data here demonstrate a valid, fraud-free pattern even when one candidate is overwhelmingly favored in the vote.

[Figures 4a and 4b here]

Now consider Figures 5a - c, which concern all precincts throughout the state of North Carolina and that cover data in which both Bush and Kerry won pluralities. Here, however, because it is safe to assume that there are marked political-economic differences between Democratic and Republican precincts (i.e., significant non-homogeneity in the data), we separate the data according to which candidate won in each precinct. Now notice (Figure 5c) the wholly unexceptional distribution of turnout among both 'Bush precincts' and 'Kerry precincts'.

Second, notice the unexceptional relationship between turnout and each candidate's absolute vote in the two "half-groups" of precincts. That is, the coefficients for both Bush and Kerry are uniformly positive and none exceed 1.0.

[Figures 5a - c here]

To illustrate the necessity of separating Bush and Kerry precincts, consider Mecklingburg county in North Carolina. If we combine all precincts, we estimate a distinctly negative relationship between Kerry's vote and turnout of the sort observed for Haritonov in Tatarstan. Specifically, we get

$$Bush_{Meck} = 1.35Turnout_{Meck} - 0.37$$

$$Kerry_{Meck} = -0.35Turnout_{Meck} + 0.37$$

However, by combining precincts here we are in fact running afoul of aggregation error. Owing to differences in their socio-economic circumstances, the precincts carried by the Democrat Kerry report on average a distinctly lower average level of turnout than those carried by the Republican Bush. In other words, our assumption of homogeneity is not satisfied and turnout correlates with a candidate's *relative* level of support. It is only when we separate Kerry precincts from Bush ones and, thereby, move in the direction of eliminating the effects of unobserved variables (i.e., socio-economic characteristics) that a normal relationship between turnout and a candidate's absolute vote appears. In fact, with Bush and Kerry precincts separated, our regressions become:

$$Bush_{Meck,Bush} = 0.84Turnout_{Meck,Bush} - 0.09$$

$$Kerry_{Meck,Bush} = 0.16Turnout_{Meck,Bush} + 0.09$$

And in the precincts Kerry carried,

$$Bush_{Meck,Kerry} = 0.66Turnout_{Meck,Kerry} - 0.14$$

$$Kerry_{Meck,Kerry} = 0.34Turnout_{Meck,Kerry} + 0.14$$

Now let us turn to several of the counties in Florida whose vote counts were controversial in the 2000 Presidential contest between President Bush and Vice President Gore. First, consider Figures 6a and b, which graph the distributions of turnout for Bush and Gore precincts in Duval and Dade counties. Once again, we find nothing here that is unusual or suspicious in these turnout distributions – an observation with respect to turnout that is repeated if we also consider other disputed counties such as Broward, and Escambia.

[Figures 6a, b here]

Much the same story is told if we next consider the relationships between Bush and Gore's vote as a share of the total eligible electorate and turnout. In Dade county we get for the precincts carried by Bush

$$Bush_{Bush,Dade} = 0.87Turnout_{Bush,Dade} - 0.15$$

$$Gore_{Bush,Dade} = 0.19Turnout_{Bush,Dade} + 0.08$$

And in the Gore precincts,

$$Bush_{Gore,Dade} = 0.32Turnout_{Gore,Dade} - 0.05$$

$$Gore_{Gore,Dade} = 0.71Turnout_{Gore,Dade} - 0.01$$

Similarly, in Duval country we get

$$Bush_{Bush,Duval} = 0.98Turnout_{Bush,Duval} - 0.26$$

$$Gore_{Bush,Duval} = 0.07Turnout_{Bush,Duval} + 0.16$$

And in the Gore precincts,

$$Bush_{Gore,Duval} = 0.76Turnout_{Gore,Duval} - 0.06$$

$$Gore_{Gore,Duval} = 0.08Turnout_{Gore,Duval} + 0.04$$

This pattern as well is repeated in Escambria and Broward counties and there is no need to report similar coefficients here.

4. Ohio, 2004

While Florida with its 'hanging chads' was the focus of attention in 2000, in the 2004 presidential contest, questions surfaced about the legitimacy of the vote count in Ohio, where in Franklin County (Columbus) the media fixated itself on a solitary precinct that reported some 4000 more votes for Bush than there were registered voters, and in Cuyahoga County (Cleveland), where Democrats alleged such tactics as too few polling stations in predominantly Democratic precincts. However, Figures 7a and b summarize the distributions of turnout for both of these counties and reveals little that is suspect.

[Figures 7a, b here]

An equivalent story is told by our third indicator, the relationship between absolute vote and turnout. Here, our regression estimates are as follows:

$$Bush_{Bush,Franklin} = 0.77Turnout_{Bush,Franklin} - 0.11$$

$$Kerry_{Bush,Franklin} = 0.22Turnout_{Bush,Franklin} + 0.10$$

And in the Kerry precincts of Franklin County,

$$Bush_{Gore,Franklin} = 0.42Turnout_{Kerry,Franklin} - 0.06$$

$$Kerry_{Kerry,Franklin} = 0.55Turnout_{Kerry,Franklin} + 0.05$$

Similarly, for Cuyahoga County:

$$\text{Bush}_{\text{Bush,Cuyahoga}} = 0.59\text{Turnout}_{\text{Bush,Cuyahoga}} - 0.02$$

$$\text{Kerry}_{\text{Bush,Cuyahoga}} = 0.41\text{Turnout}_{\text{Bush,Cuyahoga}} + 0.02$$

$$\text{Bush}_{\text{Kerry, Cuyahoga}} = 0.64\text{Turnout}_{\text{Kerry,Cuyahoga}} - 0.20$$

$$\text{Kerry}_{\text{Kerry,Cuyahoga}} = 0.36\text{Turnout}_{\text{Kerry,Cuyahoga}} + 0.20$$

To this point, then, there is noting exceptional to be found in these two Ohio counties. If voting irregularities existed, they were of too small a magnitude to be detected by our methods. Indeed, both Florida and Ohio demonstrate that our methods and their reliance on aggregate data are suited for detecting fraud on a relatively massive scale – for instances of fraud whereby, say, five percent or more of the vote is falsified or otherwise manipulated. But now as a further caution about the application of our methods, consider Figures 8a-c which summarize our two methods for Hamilton County (Cincinnati) in 2004. Here, in those precincts carried by President Bush we see a pattern in the distribution of turnout and in the relationship between share of the eligible electorate and turnout that are reminiscent of republics in Russia. The turnout distribution is bimodal, the coefficient for Kerry is significantly negative and the coefficient for Bush significantly greater than 1.0.

[Figures 8a-c here]

It might seem, then, that we have discovered the irregularities that swung an election. However, it is here that we see the necessity for applying our methods with care. First, as Figures 9a-c show, the distributions of turnout are bimodal for every presidential election in Republican precincts in Hamilton County beginning with 1994.

[Figures 9a-c here]

And second, as the following sequence of regressions document, coefficients of the sort found in

Figure 8b exist uniformly: For the 2000 election,

$$Bush_{Bush,Hamilton} = 1.05Turnout_{Bush,Hamilton} - 0.28$$

$$Gore_{Bush,Hamilton} = -0.05Turnout_{Bush,Hamilton} + 0.24$$

For the 1996 election in Dole's precincts,

$$Dole_{Dole,Hamilton} = 1.23Turnout_{Dole,Hamilton} - 0.48$$

$$Clinton_{Dole,Hamilton} = -0.17Turnout_{Dole,Hamilton} + 0.37$$

And for the 1992 election in Bush's precincts,

$$Bush_{Bush,Hamilton} = 1.14Turnout_{Bush,Hamilton} - 0.48$$

$$Clinton_{Bush,Hamilton} = -0.24Turnout_{Bush,Hamilton} + 0.40$$

There are, then, two alternative inferences to be drawn from this data: Either the Democratic party has been asleep at the switch and oblivious to significant and historically consistent fraud in Hamilton County or there is some innocuous explanation for the patterns our analysis reveals. In fact, there exists a ready explanation. Specifically, there are a number of "Republican" precincts that, owing to their demographic makeup, are closely contested and return only slight pluralities to any Republican presidential candidate. Those same precincts, though, report relatively low rates of turnout since on average, Democratic voters within them vote at a lower rate than their Republican counterparts. Were those eligible voters to vote at the same rate as their Republican counterparts, the precincts in which they reside would shift to the Democratic party column. The result is to 'add' data at the lower range of turnout, but with higher vote shares for a Democratic candidate than in other Republican precincts, thus 'lifting' the left end of any regression line for a Democratic candidate and lowering it for his Republican challenger. This, in turn, yields coefficients less than 0.0 and greater than 1.0. The existence of

such precincts, moreover, is revealed in our turnout graphs and accounts for the slight bimodal nature of the distributions there. That is, the bimodality of turnout there reveals two distinct types of districts … those that are uniformly Republican and those that are divided in their support but which yield Republican pluralities owing to relatively low turnout among Democratic party supporters.

5. Mexico 2006

Hamilton County, then, serves as a warning about the application of our methods without regard to the nature of the data being considered. Enough must be known about that data and historical voting patterns to ensure that the assumption of homogeneity is at least approximately satisfied lest a false signal of electoral irregularity be taken as an indicator of fraud. Nevertheless, the signals provided by our indicators, while not confirming fraud, can be used to direct our attention to specific potential sources of electoral irregularities. For such an example, consider Mexico's most recent and controversial presidential election in which the unsuccessful candidate, Lopez Obrador, directed charges of massive election irregularities against the official winner, President Felipe Calderon. To begin, then, consider Figure 10 which graphs the distribution of turnout for three of Mexico's election districts, one in the south carried by Obrador, one in the central part of the country carried narrowly by Calderon and one in the north carried overwhelmingly by Calderon and the focus of many of the allegations of electoral irregularity. None of these distributions matches a symmetric normal, with each seemingly having an accentuated left-most tail of districts with unusually low turnout. It is in the controversial northern district, however, that we see a somewhat muted bimodal distribution that

matches the distribution we observe for Hamilton country in Ohio.

[Figure 10 here]

Turning then to the relationship between turnout and a candidate's share of the eligible electorate, Figures 11a-c graph this relationship for each of the three regions considered in Figure 10. Notice first that the regression lines and coefficients in both the central and southern regions seems wholly unexceptional (Figures 11b and 11c). In contrast, Figure 11a offers a pair of regression lines that, once again, imitate Hamilton county. Here, of course, we need to emphasize that despite the negative coefficient for Oberon and a coefficient close to one for Calderon, Figure 11a in no way approximates what we find in Russia's ethnic republics. Hence, although suspicions are raised by this figure (in conjunction with Figure 10), there is no unambiguous evidence of massive fraud here. Surely our analysis raises suspicions, but definitive conclusions cannot be reached without further exploration of, among other things, the historical record. Just as the persistent anomalies in the data from Hamilton county suggested a more benign socio-economic explanation for things, the same may be true of Mexico. At a minimum, then, our analysis at least serves to focus our investigations on, perhaps, specific voting districts as well as, more generally, regions of a country.

[Figures 11a - c here]

Figure 1: Turnout distributions, Russia 2000 and 2004

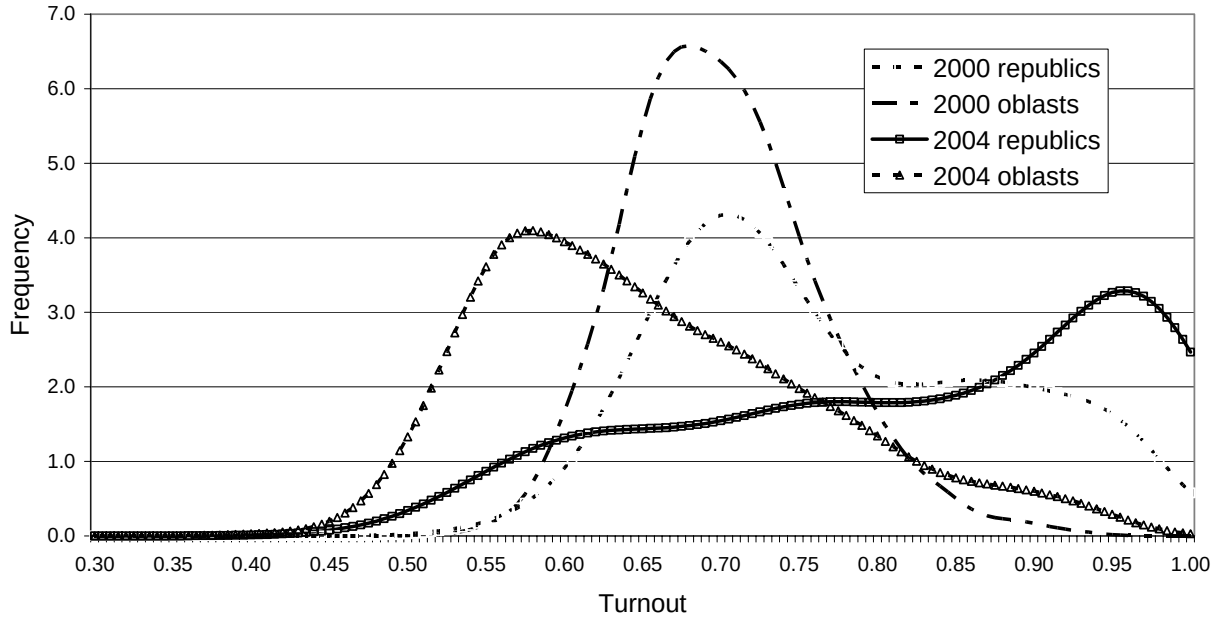


Figure 2: Turnout Distribution, Rounds 2 and 3

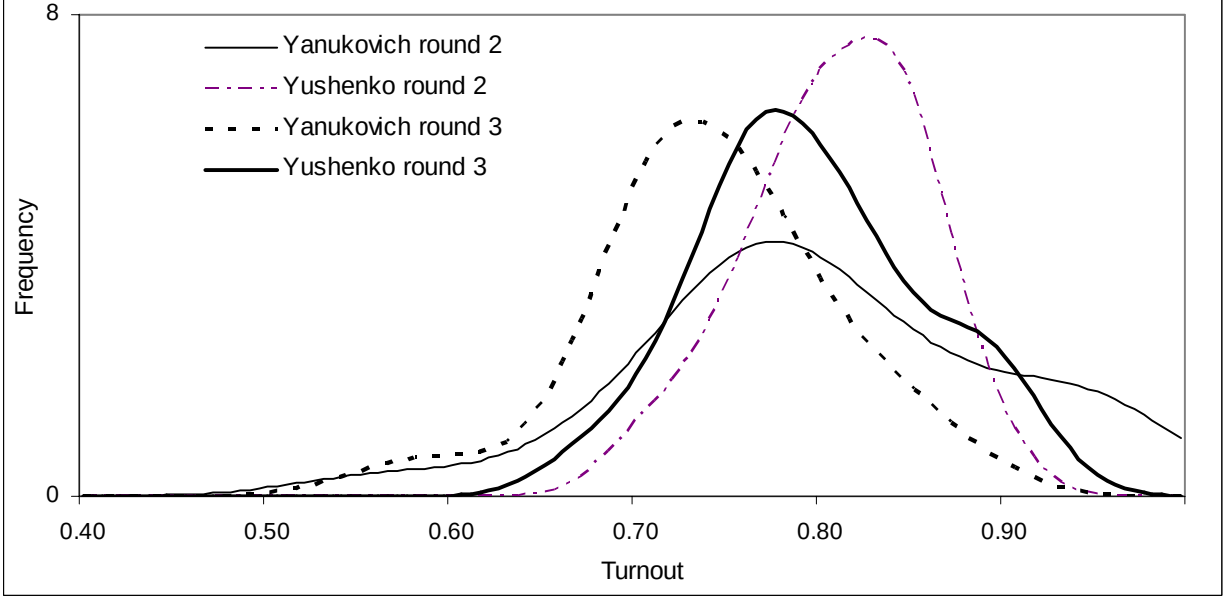


Figure 3b: Lviv, rounds 1-3

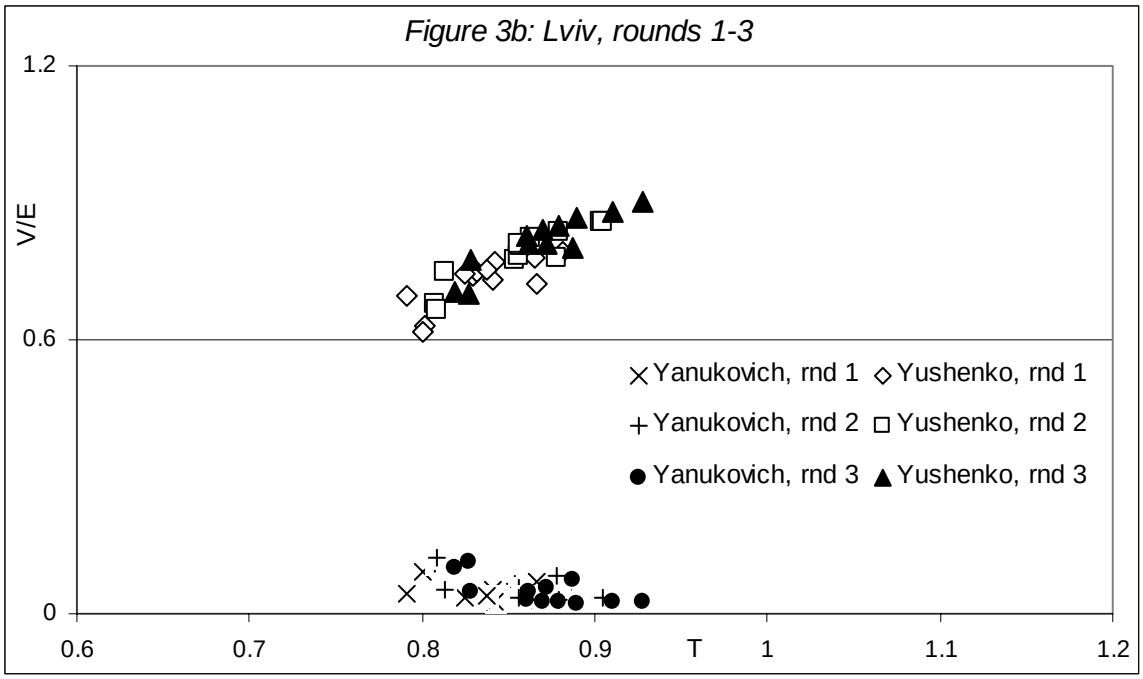
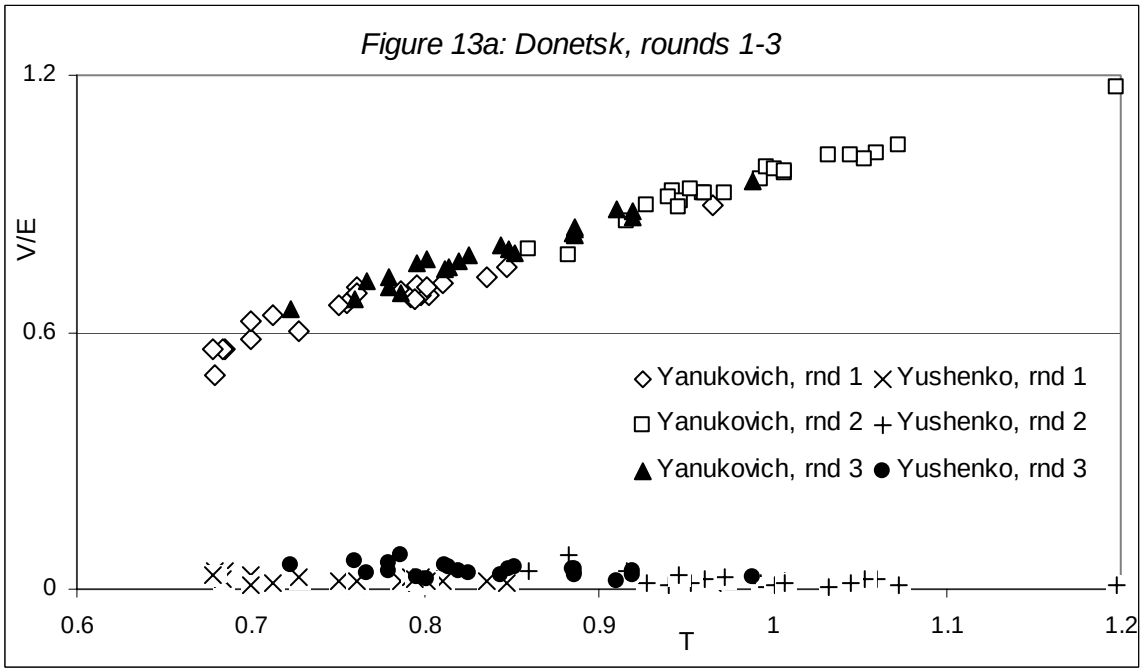


Figure 13a: Donetsk, rounds 1-3



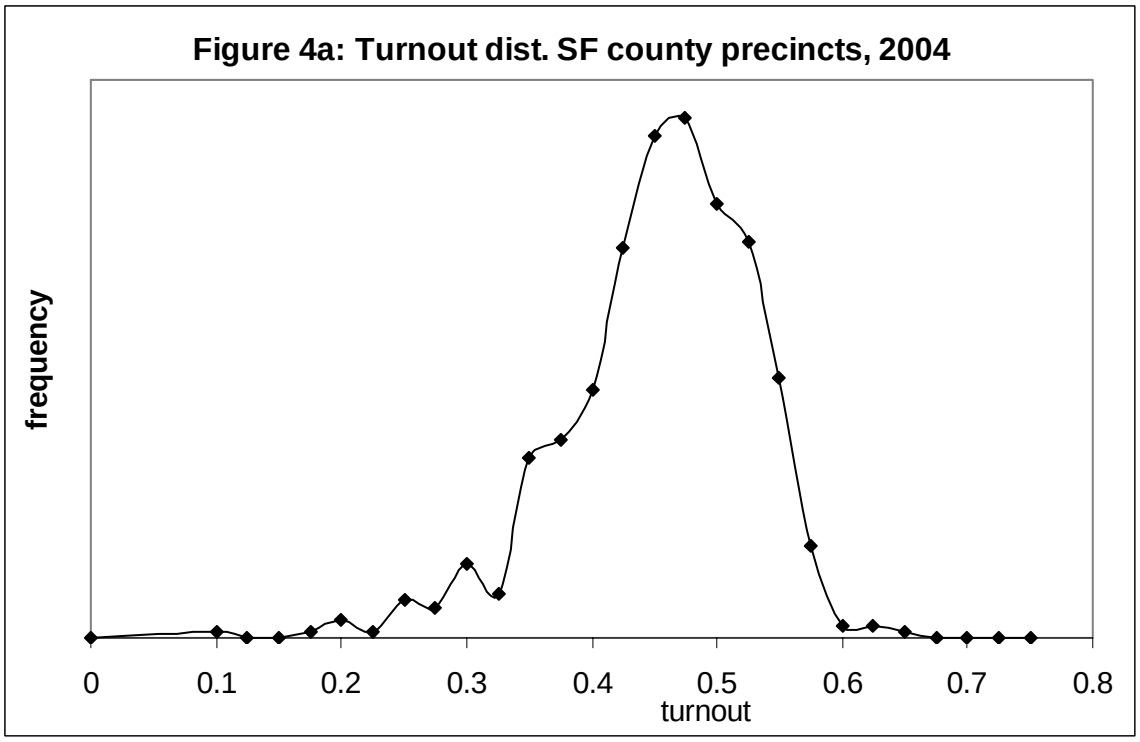


Figure 4a: Turnout dist. SF county precincts, 2004
frequency
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
turnout

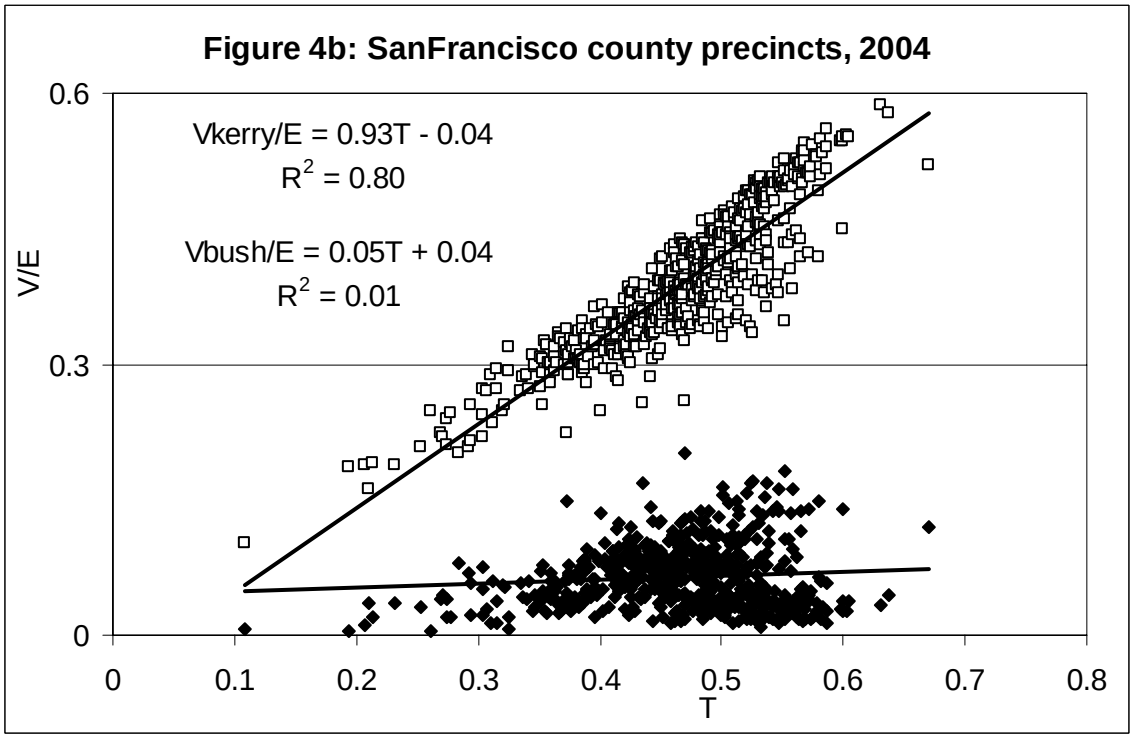


Figure 4b: SanFrancisco county precincts, 2004
Vkerry/E = 0.93T - 0.04
R2 = 0.80
Vbush/E = 0.05T + 0.04
R2 = 0.01
V/E
0.6
0.3
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
T

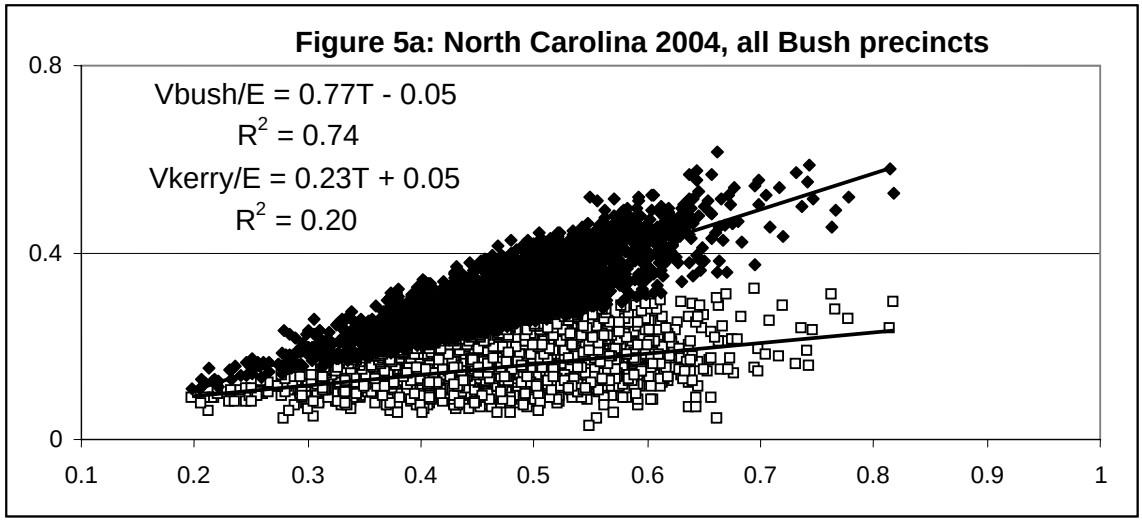


Figure 5a: North Carolina 2004, all Bush precincts
Vbush/E = 0.77T - 0.05
R2 = 0.74
Vkerry/E = 0.23T + 0.05
R2 = 0.20
0.8
0.4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

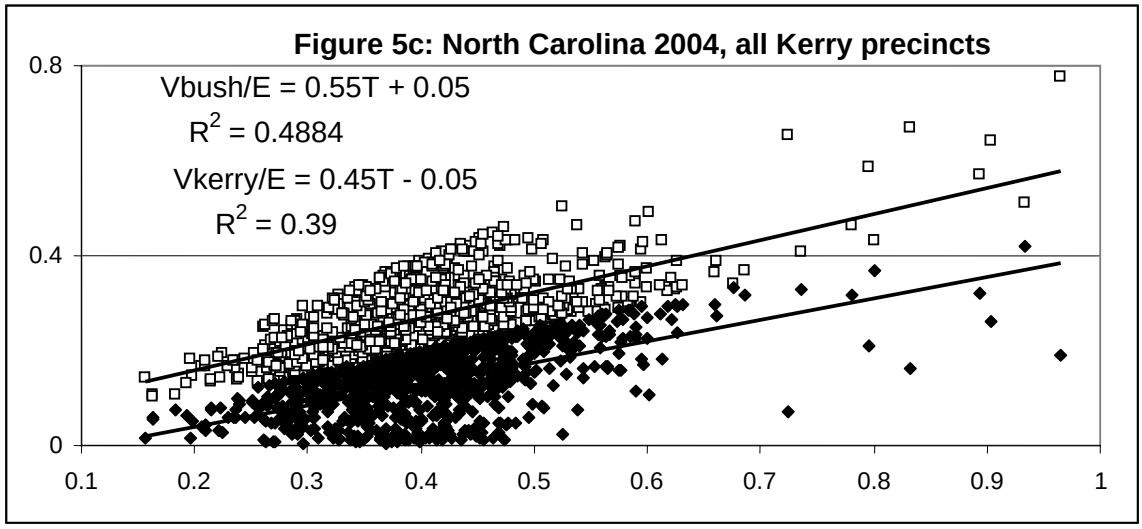


Figure 5c: North Carolina 2004, all Kerry precincts
Vbush/E = 0.55T + 0.05
R2 = 0.4884
Vkerry/E = 0.45T - 0.05
R2 = 0.39
0.8
0.4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

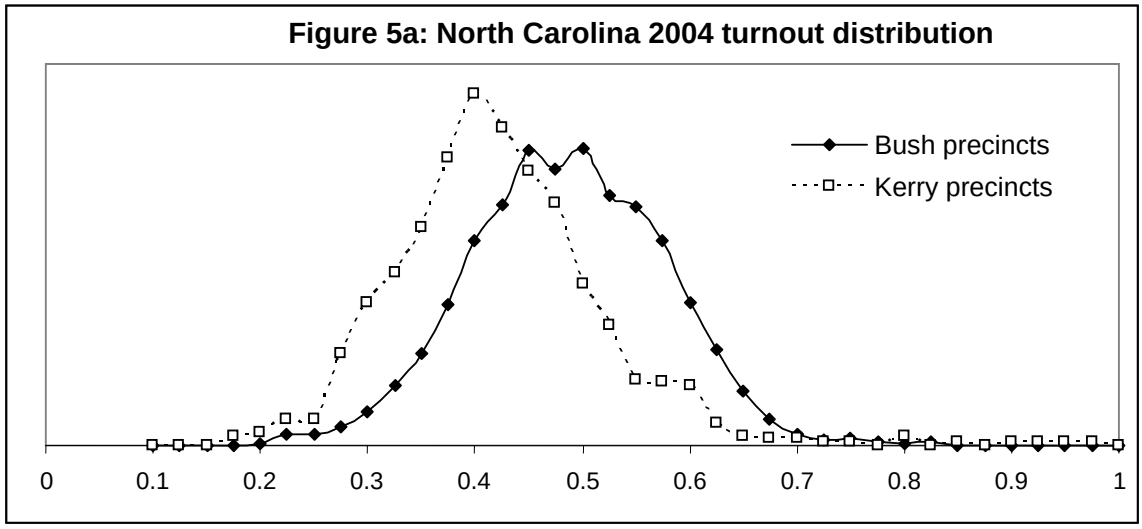


Figure 5a: North Carolina 2004 turnout distribution
Bush precincts
Kerry precincts
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

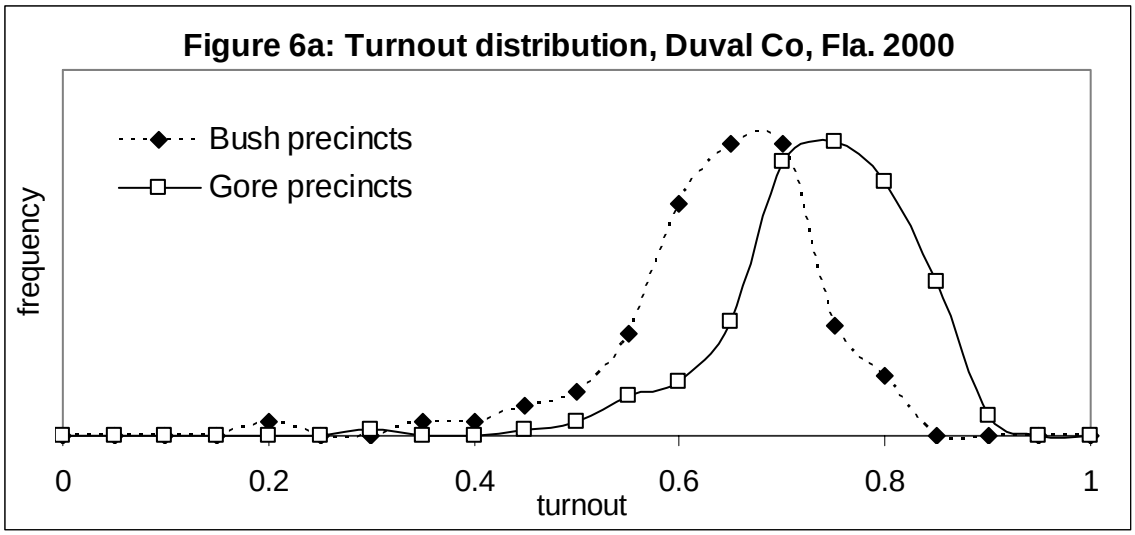


Figure 6a: Turnout distribution, Duval Co, Fla. 2000
Bush precincts
Gore precincts
frequency
0
0.2
0.4
0.6
0.8
1
turnout

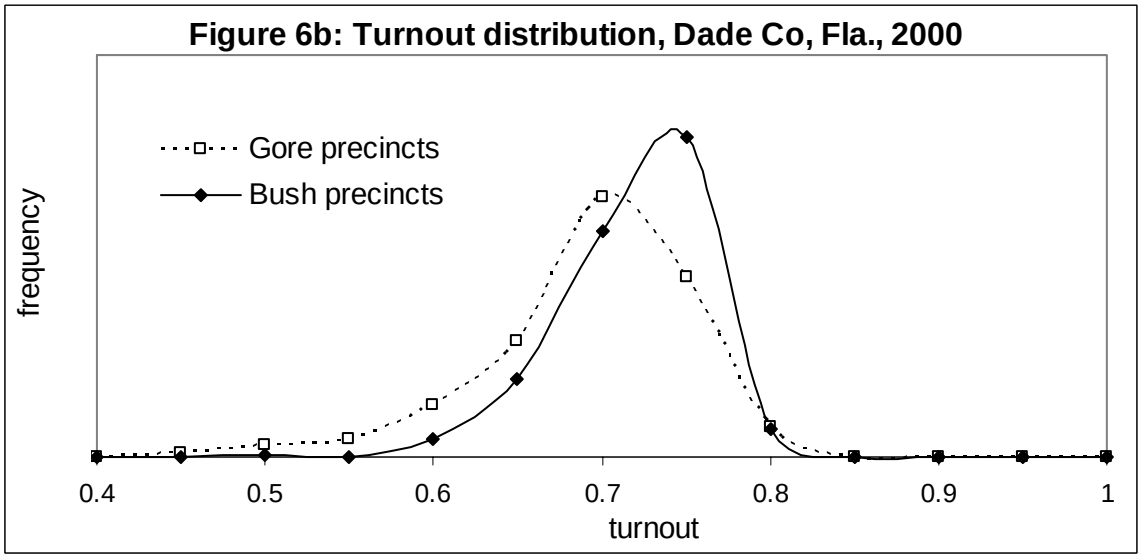


Figure 6b: Turnout distribution, Dade Co, Fla., 2000
Gore precincts
Bush precincts
frequency
0.4
0.5
0.6
0.7
0.8
0.9
1
turnout

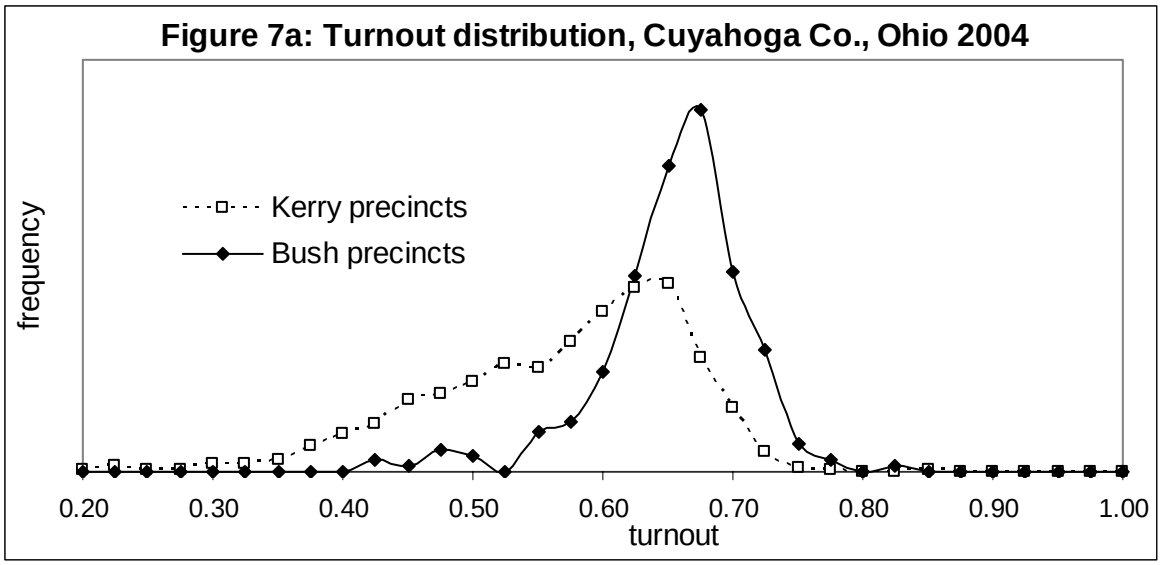


Figure 7a: Turnout distribution, Cuyahoga Co., Ohio 2004
Kerry precincts
Bush precincts
frequency
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
turnout

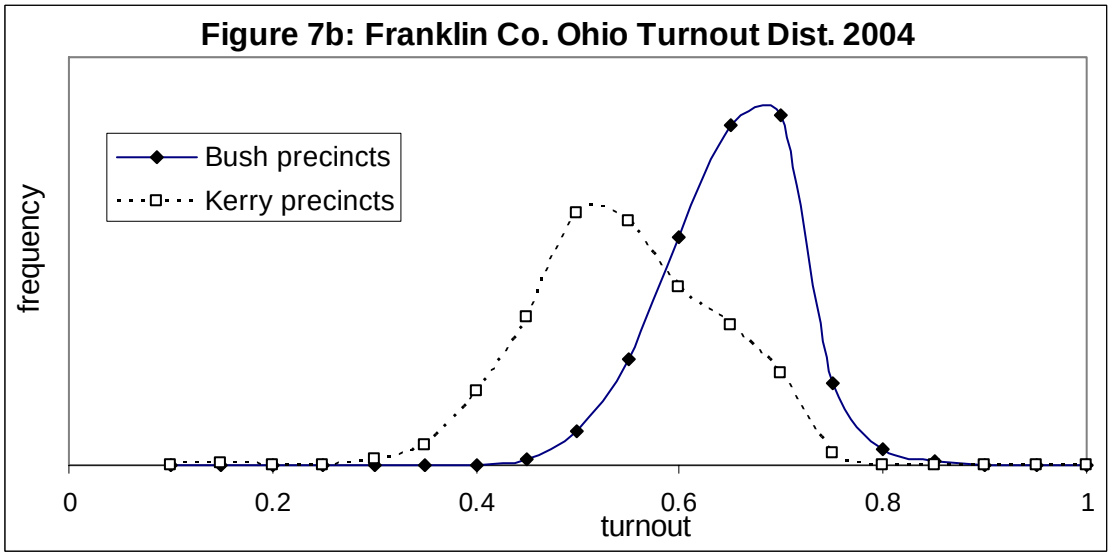


Figure 7b: Franklin Co. Ohio Turnout Dist. 2004
Bush precincts
Kerry precincts
frequency
0
0.2
0.4
0.6
0.8
1
turnout

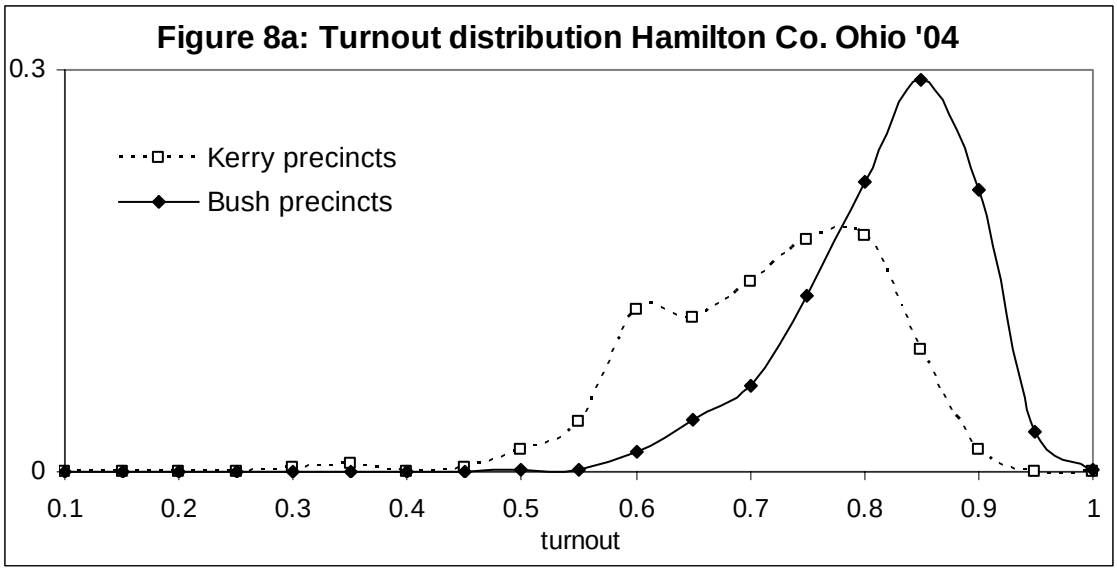


Figure 8a: Turnout distribution Hamilton Co. Ohio '04
0.3
0
Kerry precincts
Bush precincts
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
turnout

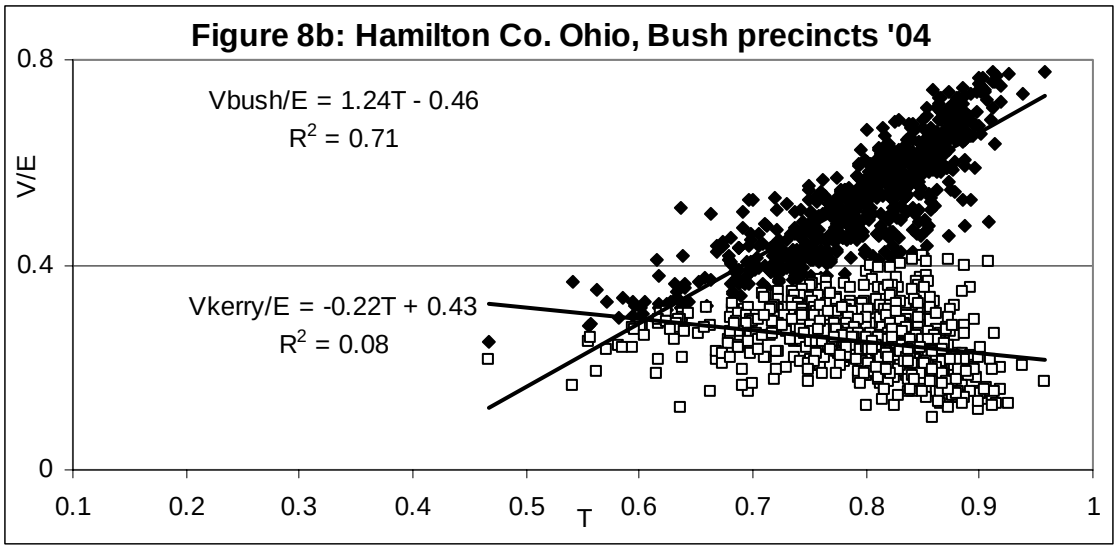


Figure 8b: Hamilton Co. Ohio, Bush precincts '04
Vbush/E = 1.24T - 0.46
R2 = 0.71
Vkerry/E = -0.22T + 0.43
R2 = 0.08
V/E
T
0.8
0.4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

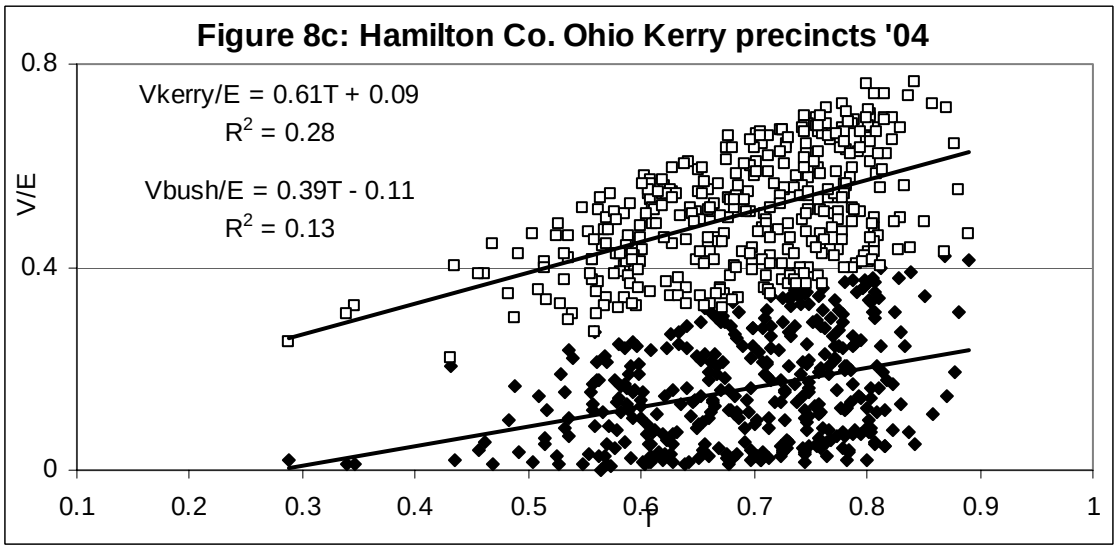


Figure 8c: Hamilton Co. Ohio Kerry precincts '04
Vkerry/E = 0.61T + 0.09
R2 = 0.28
Vbush/E = 0.39T - 0.11
R2 = 0.13
V/E
T
0.8
0.4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

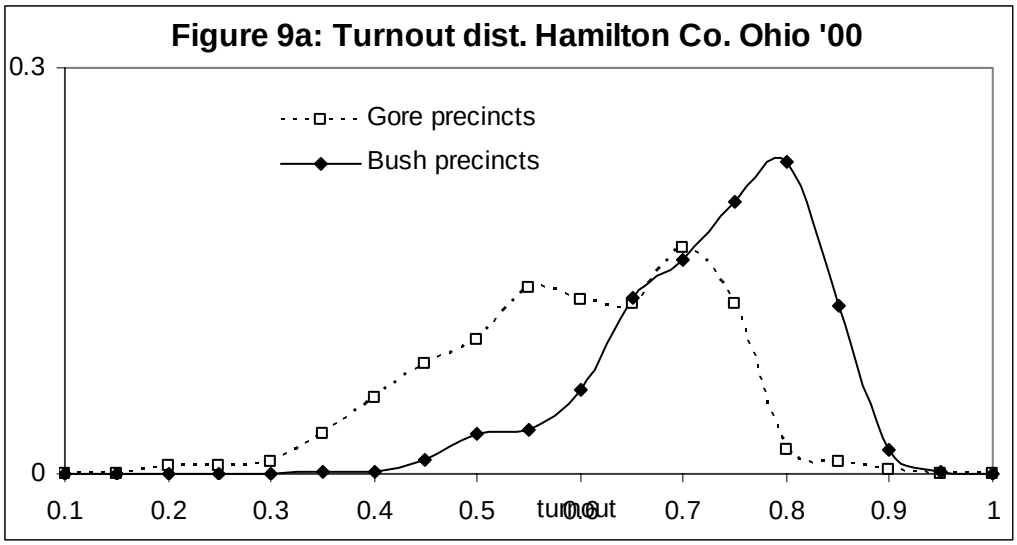


Figure 9a: Turnout dist. Hamilton Co. Ohio '00
0.3
0
Gore precincts
Bush precincts
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
turnout

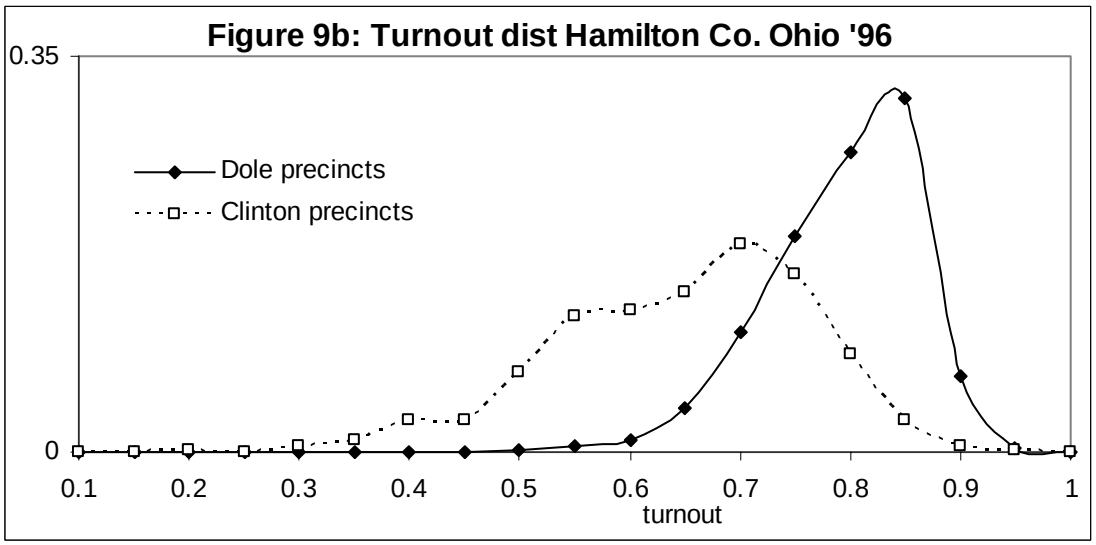


Figure 9b: Turnout dist Hamilton Co. Ohio '96
0.35
0
Dole precincts
Clinton precincts
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
turnout

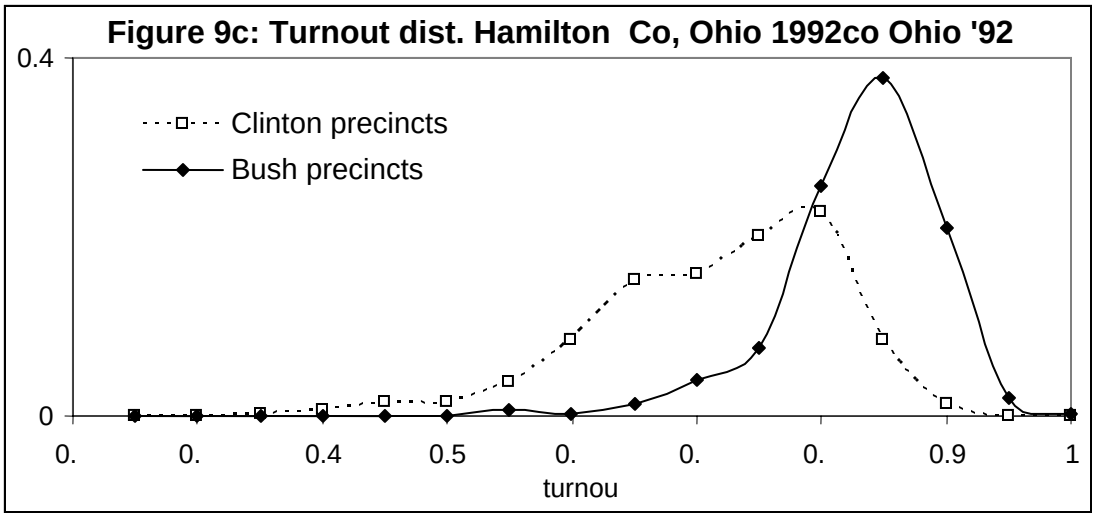


Figure 9c: Turnout dist. Hamilton Co, Ohio 1992co Ohio '92
0.4
0
Clinton precincts
Bush precincts
0.
0.
0.4
0.5
0.
0.
0.
0.9
1
turnou

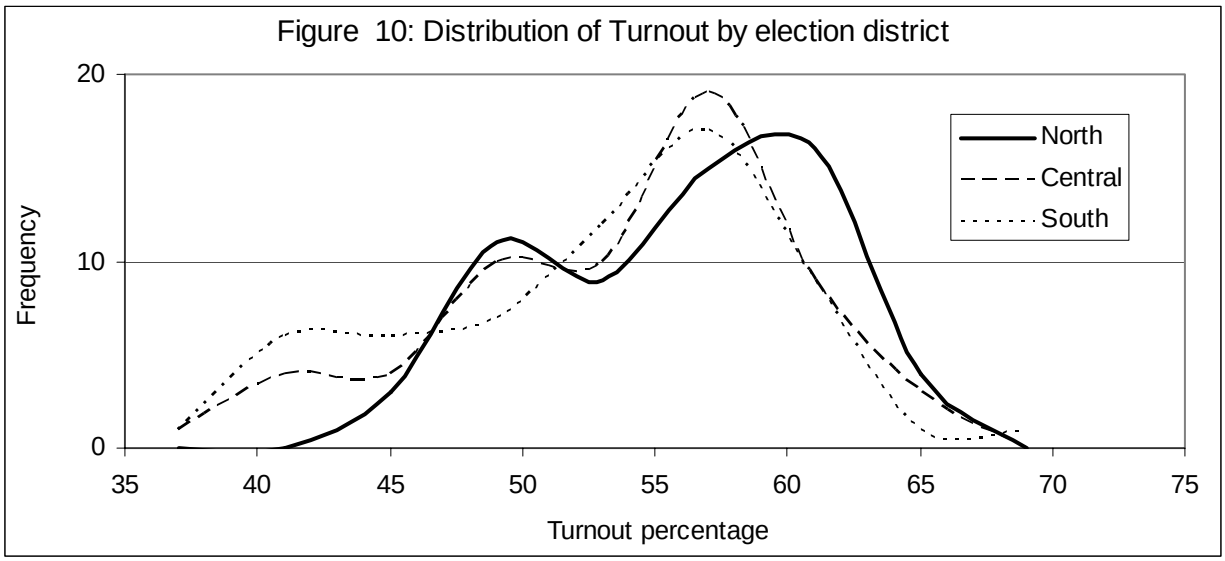


Figure 10: Distribution of Turnout by election district
20
10
0
Frequency
North
Central
South
35
40
45
50
55
60
65
70
75
Turnout percentage

Figure 11a: Northern region

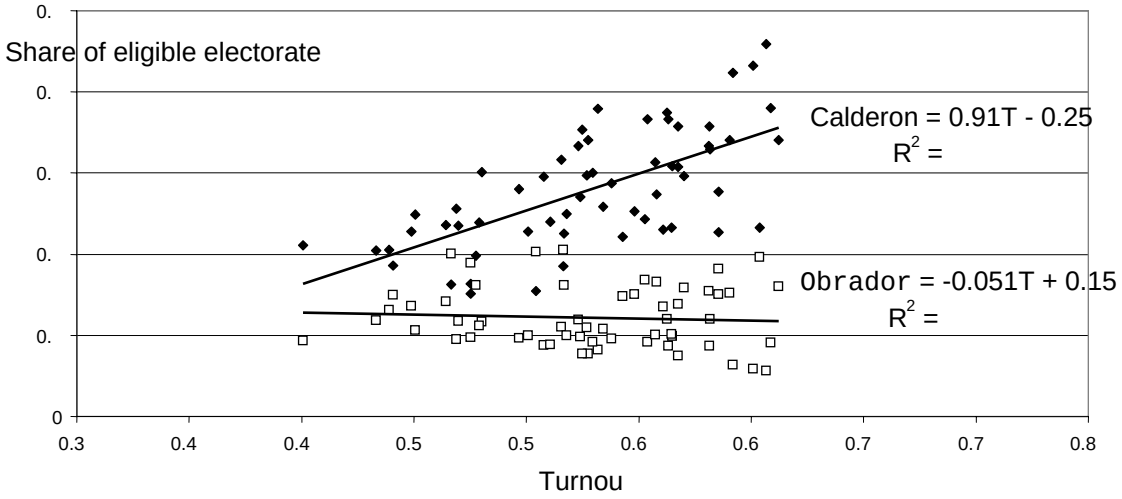


Figure 11b: Central region

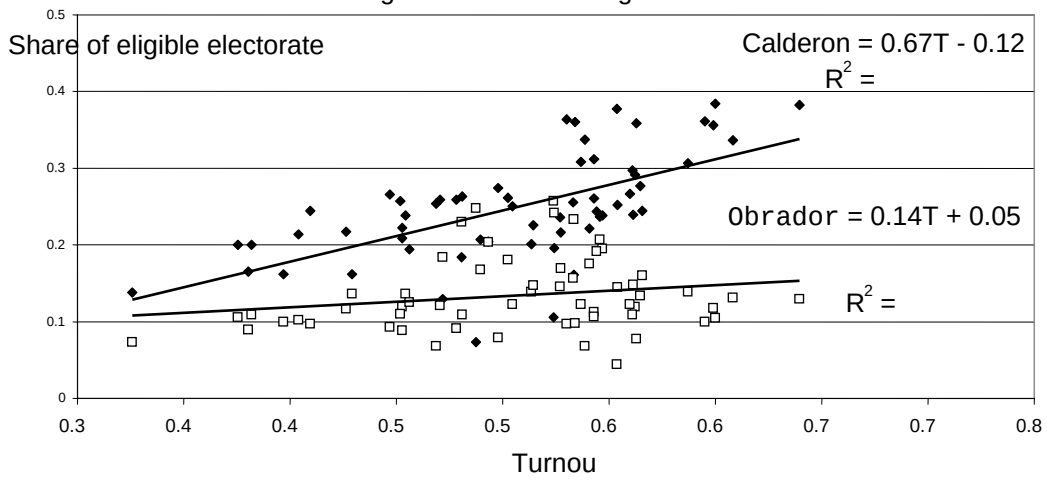


Figure 11c: Southern region