Votes lost due to under-provision of voting machines in Franklin County, Ohio

Revised Monday, January 03, 2005

USCountVotes
by Elizabeth Liddle

An article by Harvey Wasserman in The Free Press[1] alleged that long lines on Election Day, 2nd November 2004 deprived some voters in Franklin county, Ohio, of the right to vote, and that this problem was greatest in the most strongly Democratic precincts. If these allegations are well-founded, they should be apparent in the election data. I therefore analyzed data from Franklin County supplied by Cliff Arnebeck[2]. The results indicate that in precincts where the number of "active voters" (voters who have voted at least once in the last two election cycles) per voting machine was high, turnout was significantly depressed as compared with turnout in precincts where the number of "active voters" per machine was low, supporting Wasserman's claim that turnout was indeed depressed by inadequate provision of machines. The results also indicate that the number "active voters" required to share a machine in a precinct was strongly and positively correlated with the proportion of that precinct's vote for John Kerry, again supporting Wasserman's claim of selective under-provision of machines to strongly Democratic precincts. Statistical tests of both these effects indicate that they are highly significant, and thus extremely unlikely to have arisen by chance.

Two things are clear to the naked eye from the data. First is the finding that where the number of active voters per machine in a precinct exceeded a certain level, turnout (as a proportion of potential voters) dropped off (Figure 1).

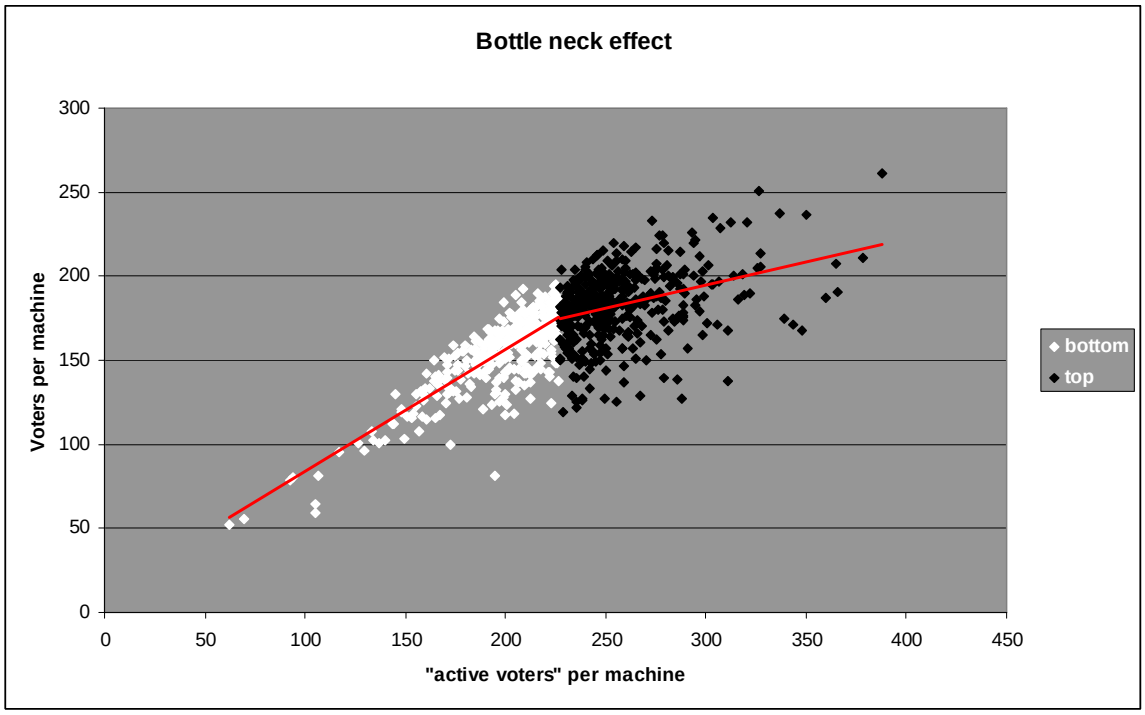

Figure1: Each dot represents one precinct. The horizontal axis represents the number of active voters/machine for that precinct, and the vertical axis represents the number of actual voters/machine. The white data points are precincts whose number of active voters/machine was less than the median of 227. The black data points represent precincts with more

[1] November 17th, 2004 at: http://www.freepress.org/columns/display/7/2004/991
[2] available at http://copperas.com/machinery

than 227 “active voters” per precinct. The cranked red line represents the best fit (regression line) through each half of the data.

Two methods were used to test the hypothesis that the slope was significantly flatter at the right hand end of the graph than at the left hand end. The first method simply divided the precincts at the median level of 227 active voters/precinct, and compared the slopes in each half of the data. The second tested whether the best fit line through the whole data set was significantly non-linear. Both statistical tests indicated that the observed flattening of the slope at high levels of active voters/machine is highly unlikely to have arisen by chance.[1]

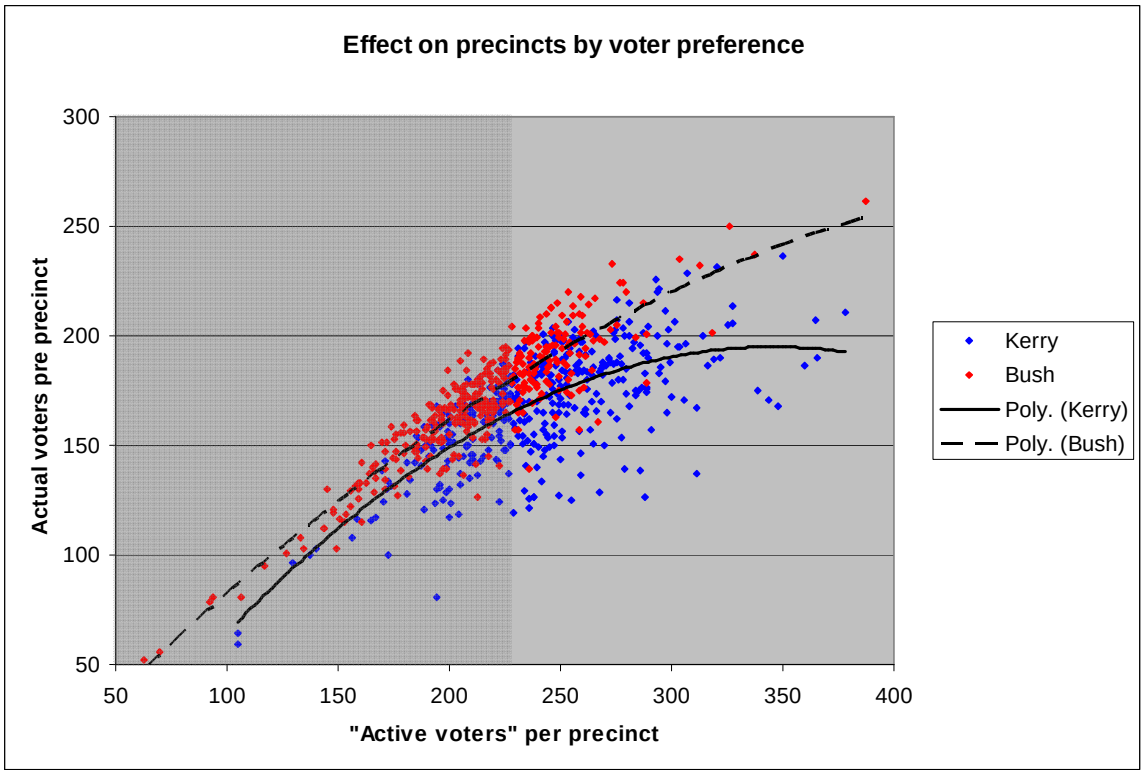

Figure 2: Data plotted to show precincts in which Kerry won less than 50% of the vote (red) and those in which he won more than 50% (blue). The darker background indicates precincts below the median of 227 active voters/machine, Note that the Kerry precincts are shifted to the right hand end of the graph, indicating that more Kerry precincts were above the median level of 227 active voters/machine. It is also apparent from this graph that turnout is lower in the pro-Kerry precincts. Therefore, to demonstrate that this is not the sole reason for the shallower slope in the right hand end of the graph, I have plotted quadratic polynomial fit lines for each group of precincts. The lines are parallel at the left of the graph, but diverge on the right, with a greater flattening in the Kerry precincts. A possible explanation for this may be that, with a complex ballot, educational and mobility factors affecting the time taken to vote impacted more strongly on the pro-Kerry precincts.

[1] Both methods used the General Linear Model. In the first, the precincts were divided into two groups at the median. The null hypothesis was that there would be no significant interaction between *active voters/machine* and *group* in predicting *voters/machine*. The interaction term was highly significant [$F(1,784)=96{,}682$, $p<1E\text{-}21$), indicating a significant difference between the slopes. However, because dividing at the median involves a violation of homoscedasticity assumptions, an alternative approach tested whether, across the whole sample, the addition of a non-linear function of active *voters/machine* would result in a significant increase in variance in *voters/machine* over and above the variance accounted for by a linear function of active *voters/machine*. *Voters/machine* was therefore regressed on a quadratic function of *active voters/machine*, the quadratic term being entered second into the regression. Entry of the quadratic term resulted in a significant increase in R^2 [$F(1, 785)=94.182$, $p<4E\text{-}21$]. The signs and values of the regression coefficients indicated a decrease in slope at increased values of active voter per machine.

The most straightforward interpretation of this finding would seem to be that as the number of active voters/machine rose, turnout started to decline. This suggests that under-provision of machines resulted in a significant proportion of would-be voters being unable to vote.

Secondly, it is apparent (Figure 2) that precincts in pro-Kerry precincts were most likely to lie at the right hand end of the graph, where active voter/machine numbers are above the median. To test this effect for statistical significance, I used a non-parametric technique (which makes no assumptions about the pattern of distribution in the data) to determine the correlation between the proportion of Kerry voters in a precinct and the number of active voters/machine. This correlation is positive and highly significant[1]. It also remained robust when precincts with the minimum of two machines (which might unfairly bias the correlation) were excluded.

In other words, the naked eye rightward shift in the "Kerry" precincts is a statistically significant effect, and seems inexplicable: why should the way you vote affect how many active voters have to share a machine in your precinct? Further statistical analyses may shed some light on the question.

I do not know the formula used by the Board of Elections to allocate machines, except that there appears to have been a minimum of two machines allocated to each precinct. One possibility is that allocation was done using active voter figures for each precinct. If so, and assuming a minimum of two machines per precinct, I found (using a trial-and-error procedure) that the formula that best fitted the actual data (numbers of machines actually provided in each precinct) was one that allowed for a maximum of 250 active voters per machine. This formula exactly accounts for machine totals in 55% of precincts, with 21% of precincts getting one more machine than predicted by the formula, and 13% getting one less. It thus accounts for 76% of the data within one machine and 89% within two. Precincts which got more machines than predicted by this formula had a lower proportion of Kerry voters; precincts which got fewer had more Kerry voters. If this was the formula used, there would appear to have been a deliberate withholding of machines from pro-Kerry precincts, and a re-allocation to pro-Bush precincts. If so, it would seem to be tantamount to deliberate suppression of the Kerry vote.

However, there is another possibility. Given that not enough machines appear to have been purchased (for unknown reasons), it would have made sense to take turnout in 2000 into account when allocating the machines. Using a hypothetical allocation formula based on turnout in 2000 only, I found that the best fit to the data was obtained by a formula that again allocated at least two machines per precinct, and a maximum of 180 year 2000 voters per machine. This formula accounts for 62% of precinct allocations. According to this formula, a further 23% of precincts were allocated one extra machine. The formula thus accounts for 89% of the data within one machine, and thus fits the data better than the "active voter" formula, despite 2000 turnout data being unavailable for 4% of precincts, Precincts allocated an extra machine (defined by this formula) were *more* likely to vote for Kerry than those that were not.

Why, then, if it was used, did it result in the "Kerry" precincts having so significantly fewer machines per active voter? The reason would appear to lie in the fact that turnout in 2000 was significantly[2] lower in precincts which, in 2004, voted most strongly for Kerry. In other words, whether a "turnout in 2000" formula was weighted by "active voter 2004" numbers, or whether an "active voter" formula was weighted by "turnout in 2000", the end result was a net deprivation of machines to pro-Kerry precincts.

As turnout in 2000 was significantly negatively correlated with Kerry's percentage of the vote in 2004 (and therefore, it can be inferred, with habitual party allegiance), any weighting of machine allocation by turnout in 2000 is tantamount to weighting by party preference. Because the sheer supply of machines was inadequate, this was likely to result in a cap on turnout in pro-Kerry precincts at 2000 levels.

I therefore attempted to quantify the effect in terms of numbers of lost votes.

[1] Spearman's rho= 0.36, p<1E-6

[2] Correlation between Kerry's percentage of the vote in 2004 and turnout in 2000 as a percentage of 1) active voters (2004): Spearman's rho = -0.447, p<.1E-6; as a percentage of registered voters in 2000: Spearman's rho - 0.572, p<.1E-6

Although it is clearly impossible to arrive at a definite number, an estimate can be made if we assume that below the median level of 227 active voters/machine, crowding did not affect turnout, and then extrapolate from this half of the data into the half of the graph where crowding did appear to affect turnout. It will also be important to factor in any difference in turnout that is characteristic of Kerry voters versus Bush voters, particularly as Figure 2 suggests that even where machine allocation was relatively generous, turnout in 2004, as in 2000, was lower in Kerry precincts. I therefore took the precincts in which the number of machines was relatively generous (precincts with less than 227 active voters/machine), and ran a multiple regression model, in which *voters/machine* was regressed on *active voters/machine* and *Kerry percentage of the vote*. Both predictors were statistically significant; not only was there a significant positive linear slope (as shown in Figure 1, white data points) between *active voters/machine* and *voters/machine*, but *Kerry percentage of the vote* was also a significant negative predictor[1] (Figure 2: blue data points in the dark portion of the graph are lower than the red data points). Precincts with greater Kerry support had a significantly lower turnout (i.e. fewer actual voters/machine for a given number of active voters/machine. It is therefore important to model this behavior when predicting "lost votes" in the upper part of the graph.

I therefore calculated, using the regression equation[2] generated by the first model, the predicted number of actual voters per machine for each precinct across the whole county. This allowed me to predict the number of votes had there been no "bottleneck" effect. I then calculated the difference between the numbers of votes predicted by the model, and the actual numbers of votes. This gives me an estimate of votes lost through the bottleneck effect. Finally, using the proportion of votes cast for Kerry and Bush respectively for each precinct, I calculated how many of these would have been Kerry votes, and how would have been for Bush.

On these assumptions, the calculation indicates that a total of around 18,500 votes were lost. Of these, about 66% would have been cast for Kerry, and 30% for Bush; the remainder would have been cast for independents. This gives a net loss to Kerry of about 7000 votes in terms of Bush's final margin. Note that this model factors in the lower turnout expected from Kerry voters.

However, this estimate does ignore a further potential confound, namely, the relationship between the number of machines allocated per active voter, and turnout in 2000.

Because the number of machines allocated to each precinct appears to have been strongly weighted by 2000 turnout levels in that precinct, the number of "active voters" allocated to each machine is significantly negatively correlated with turnout in 2000 ($p<1E-21$). In other words, the lower turnout was in 2000, the fewer machines were allocated to a precinct in 2004. Secondly, turnout in 2000 is positively correlated with turnout in 2004 ($p<1E-302$). This means that turnout in 2000 should be controlled for when assessing the impact of machine provision *per se* on turnout in 2004.

To do this, I regressed the dependent variable, "turnout in 2004" on two predictors: "turnout in 2000" and "active voters per machine". Turnout was expressed in votes as a proportion of registered voters for each precinct. As expected, "turnout in 2000" was a significant positive predictor ($p<1E-150$) of "turnout in 2004". However, even after allowing for the highly significant linear relationship between turnout in the two years, there was an additional significant negative linear relationship between the number of "active voters" allocated to each machine ($p<.001$) in a precinct, and turnout in that precinct. In other words, even after controlling for turnout in 2000, where the number of "active voters" required to share a machine was large, turnout in 2004 was depressed.

Wassermann's allegation that under-provision of machines resulted in lost votes therefore remains supported even after allowing for the confounding relationship between the number of "active voters" allocated per machine in a precinct and that precinct's turnout in 2000.

Although the significant correlations between "active voters" per machine and turnout in 2000 means that it is more difficult to estimate either the absolute numbers of disenfranchised voters, or the proportions by which they would have voted for each presidential candidate, Wasserman's second allegation, that

[1] Regression equation: *voters/machine* = 27.38 + 0.77 x *active voters/machine* - 52.37 x *Kerry percent of vote*. All regression coefficients significant at p<3E-01 or lower.

machine shortages were greatest in pro-Kerry precincts, also remains supported, whether or not this occurred as a by-product of a strategy of weighting machine allocation by turnout in 2000. The data remain consistent with the hypothesis that an overall shortage of machines served as a "cap" on turnout in precincts with historically low-turnout, and that because voters in these precincts tended to vote predominantly for Kerry-Edwards, a disproportionate number of potential Kerry-Edwards voters failed to cast their vote this year in Franklin County.

Conclusion

The data available for Franklin county support both Wasserman's allegations:

- Under-provision of machines affected turnout
- The more strongly Democratic the precinct, the more "active voters" had to share a voting machine.

As a consequence, there would appear to have been substantial disenfranchisement of voters wishing to vote for either candidate, but particularly of those who would have liked to vote for Kerry.

Elizabeth Liddle