A Statistical Note on the April 14 Venezuelan Presidential Election and Audit of Results

David Rosnick and Mark Weisbrot

May 2013

Center for Economic and Policy Research
1611 Connecticut Avenue, NW, Suite 400
Washington, D.C. 20009
202-293-5380
www.cepr.net

Contents

Acknowledgements

The authors would like to Dan Beeton and Sara Kozameh for editing and helpful comments.

About the Author

Mark Weisbrot is an economist and Co-director of the Center for Economic and Policy Research in Washington D.C., and David Rosnick is an economist at the Center for Economic and Policy Research.

Introduction and Main Results

The results of Venezuela's April 14 presidential election returned 7,575,506 votes for Vice President and acting president Nicolás Maduro, and 7,302,641 votes for challenger Henrique Capriles Radonski. This is a difference of 272,865 votes, or 1.8 percent of the two-way total between the candidates.[1]

Capriles has challenged the result, and has asked for a "recount." The CNE (National Electoral Commission) which oversees elections, has agreed to an audit of the approximately 47 percent of machines that were not already audited on the day of the vote.[2]

In this election, voters expressed their preference by pressing a computer touch-screen, which then printed out a paper receipt of their vote. The voter then checked to make sure that the receipt was the same as her choice, and deposited the paper receipt in a covered box.

When the polls closed, some number of machines at each voting center were sampled-- some 53 percent of all the machines (20,825 out of 39,303). For each such machine, a manual tally was made of the paper receipts. This "hot audit" was done on site, in the presence of the observers from both campaigns, as well as witnesses from the community.

The opposition had asked for the same audit to be done of the remaining 46 percent of voting machines. It is not clear that this will be done, or even if it is possible. In Venezuela, the legal vote is the machine vote (as in the United States where electronic voting is used). The hot audit of 53 percent of the machines is used to ensure that there were no problems with the machine vote count. In that sense the word "recount" is misleading – the paper receipts are not the legal measure of the votes that have been cast.

Nonetheless, it may be instructive to estimate the probability that such an audit, if it were done, could add to our understanding of the election, given the results of the first audit of 53 percent of the machines. The following analysis will estimate an upper bound for that probability. There are a number of ways to do this, but as one might imagine, it will not make any practical difference. No matter how the statistical analysis is done, the probability of getting the audit results obtained on April 14th, if the remaining machines actually contained enough errors to change the result of the election, are infinitesimally small – less than one in many, many trillions.

To be sure, a difference of 272,865 votes—only 1.8 percent of the two-way vote total—is close. However, it is still relatively large in comparison to other contested elections such as Felipe Calderón's 0.8 percent two-way vote margin over Andrés Manuel López Obrador in 2006 (0.6 percent of the total vote). In that electoral result, which was based on a hand count of paper ballots, about half of the ballot boxes had "adding up" discrepancies – in other words, the number of left-over blank ballots plus used ballots did not add up to the initial amount of blank ballots, as they are supposed to do.[3] The Mexican electoral authorities did a partial recount of 9 percent of the ballots,

1 Last week, the results of voting by Venezuelan citizens in other countries were incorporated into the totals, changing the result to 50.61 percent for Maduro and 49.12 percent for Capriles. However, this change does not significantly affect the analysis of this paper.

2 There was also audit on the day following the election of 1.02 percent of the machines, randomly selected.

3 Weisbrot, Sandoval and Paredes-Drouet (2006).

but refused to release the results.[4] (It later turned out that this recounted sample was significantly different from the overall total, in favor of López Obrador.)[5]

In this election, by contrast, there was an on-site audit to determine whether the machine count matched the paper receipts. It is therefore possible to calculate the probability of getting the audit result of April 14, if in fact the unaudited machines contained errors sufficiently numerous to reverse the results of the election.

To start, it is important to note that the audit came back clean. That is, there are no reports of discrepancies between the audited machines and the hand count of the corresponding paper receipts. However, the sample was so large that even if the first audit had uncovered a number of discrepancies that somehow went unreported by the numerous witnesses, the present analysis would not change. In other words, even with, for example, a dozen discrepancies in the first audit, the calculation below would still show that the probability of getting the first audit, if the unaudited machines could change the result, is not significantly different from the results calculated below.

If we start with the audit results that were found, we can ask what is the probability that an audit of 53 percent of machines found no discrepancies, when there were in truth K discrepancies in the entire population of machines, for various levels of K. (See appendix A for more detail.) For example, if there were in truth no misreported machines in the entire population (including the 46 percent not audited), then the probability of finding zero discrepancies in the "hot audit" would be one. But this probability falls very fast as we assume more discrepancies among the total number of machines. If we assume that there were 10 discrepancies, finding none would have a probability of one in 1900. If there were 50 discrepancies to be found, and the hot audit found none, that is a result that would only happen less than one in 25 thousand trillion times. And as we will see, there would need to be quite a bit more than 50 discrepancies in the remaining machines in order to change the result of the election. We can therefore conclude that the audit that took place on April 14, of 53 percent of the machines, was decisive; and that the chance of getting the April 14 audit results, which confirmed the election results, if in fact the audit of the remaining machines would give the election to Capriles, are so infinitesimally small as to be beyond consideration. (See Appendix for calculation.)

Estimating the Probability of an Erroneous Result

To examine what an erroneous election result, given the April 14 audit, would mean, consider how many discrepancies we would need to find in order to reverse the result in an audit of the remaining machines. If there were 50 discrepancies, and each of these had favored Maduro, giving him 1,500 extra votes at each machine – which is an impossible swing in each instance – even that would not be enough to give Capriles enough votes to win the election. He would net only 75,000 votes from reversing these errors. And this would involve an event – i.e. the audit that was done -- that has a probability of less than one in 25 thousand trillion.

4 Center for Economic and Policy Research (2006).
5 Center for Economic and Policy Research (2006a).

To put this in some sort of context, imagine winning the grand prize in the Powerball lottery. Your odds are 1 in 175,223,510. Your odds of winning *both* Powerball jackpots in a single week are about 1 in 30,000 trillion.

How few could be the number of errors and how large would the errors have to be in order to throw the election into doubt? Let us suppose a particular form of misreporting error. Specifically, let us assume that instead of the reported result at any particular machine, the "true" result is 100 percent turnout all voting for Capriles. The biggest potential net pickup of votes is 1043 (574 potential votes, with Maduro recorded as the winner at this particular machine by 469 votes)). The next largest potential pickup is 1,042 (571 potential votes, and Maduro recorded as winner by 471)). If the top 313 machines are in error in this extreme way, then Capriles could eke out a 119-vote win.

But this would require a swing of 832-1043 votes per machine, in these machines, favoring Capriles. It is extremely hard to believe that errors of this magnitude went unnoticed during the audit. Instead, we would have to believe that these machines were not audited. Of course, 96 of these 313 were at voting centers with only one machine and therefore must have been audited.

If n machines were audited at a voting center, we must eliminate from consideration the n machines with the least potential swing from Maduro to Capriles. Thus, Capriles would need 325 errors in 216 voting centers, requiring swings of at least 797 net votes per machine. If this theory were held to be true, it would demand that those voting centers have audited machines showing Maduro as a heavy favorite and unaudited machines as having not only 100 percent turnout, but unanimous support for Capriles. Even ignoring the extremely low probability of such a wide difference in results at any given voting center, the odds of the selection process missing all 325 of these machines[6] would be less than 1 in 10^{106}.

It is important to note that the results in this paper do not depend on the 53 percent sample in the first audit being representative of the total population of machines. As it turns out, the audited sample was biased towards Maduro. This is because the selection process required that when there is only one machine in a polling place, that machine was audited. If there were two machines, then both were audited; three to six machines, two were audited; and larger polling places with more machines also had less than half audited. This over-represents smaller polling centers, and apparently Maduro performed relatively better in the smaller polling places.

However, the above makes no difference to this statistical analysis. This is because we are not looking at the vote total in the first audit, and doing a statistical test for the probability that we would get this vote total, given that the true vote is as recorded for all the machines (50.8 – 49.0 percent). In that case, the bias of the sample could be important. However this statistical test is simpler. It is using the result of the first audit to estimate the probability that the remaining, non-audited, machines could have misreported the actual vote by a margin large enough to reverse the result. In this analysis, what matters is how many discrepancies there are between the machines and the paper receipts in the audited sample – not whether the audited sample is representative of the total population of voting machines.

6 This is akin to performing separate audits of the 13,796 voting centers. In the 13,580 centers where there are assumed no discrepancies, the local audit must come back clean. In the remaining 216 centers, we must get unlucky in every one in order to fail to sample the machines with errors.

Appendix: Statistics of Large Audits

Suppose we audit a pool of n machines selected at random out of the entire population of N. There are

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

possible ways to select the audit sample.

Suppose the population of N machines contains K errors. If our sample happens to include k machines that are in error and $n-k$ which are not, then there are

$$\binom{K}{k}\binom{N-K}{n-k}$$

ways in which that might happen. Thus, the probability of finding exactly k errors in the audit of n machines is given by

$$p(k;K) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Thus, an entirely clean audit $(k=0)$ has probability

$$p(0;K) = \frac{\binom{K}{0}\binom{N-K}{n}}{\binom{N}{n}} = \frac{\frac{K!}{0!\,K!}\frac{(N-K)!}{n!\,(N-K-n)!}}{\frac{N!}{n!\,(N-n)!}} = \frac{(N-K)!}{N!}\frac{(N-n)!}{(N-K-n)!}$$

Note that if $n=0$, then $p(0;K)=1$. That is, if there is *no* audit, it is guaranteed to find zero errors. Likewise, if $K=0$, then $p(0;K)=1$. That is, if there are no errors, then the audit again is guaranteed to find none.

Also, if we audit a sufficient number of machines (more than $N-K$) then we are guaranteed to find any errors if there are any. This is because there are only $N-K$ machines without error, so there is no way to select $N-K+1$ machines without including at least one error.

If $0 < n \le N-K < N$, however,

$$p(0;K) = \frac{(N-n)(N-n-1)\cdots(N-n-K+1)}{(N)(N-1)\cdots(N-K+1)}$$

Note that increasing K by 1 decreases the probability by a factor of $\frac{N-K}{N-n-K} > \frac{N}{N-n}$, which means that

$$p(0;K) < \left(1-\frac{n}{N}\right)^K$$

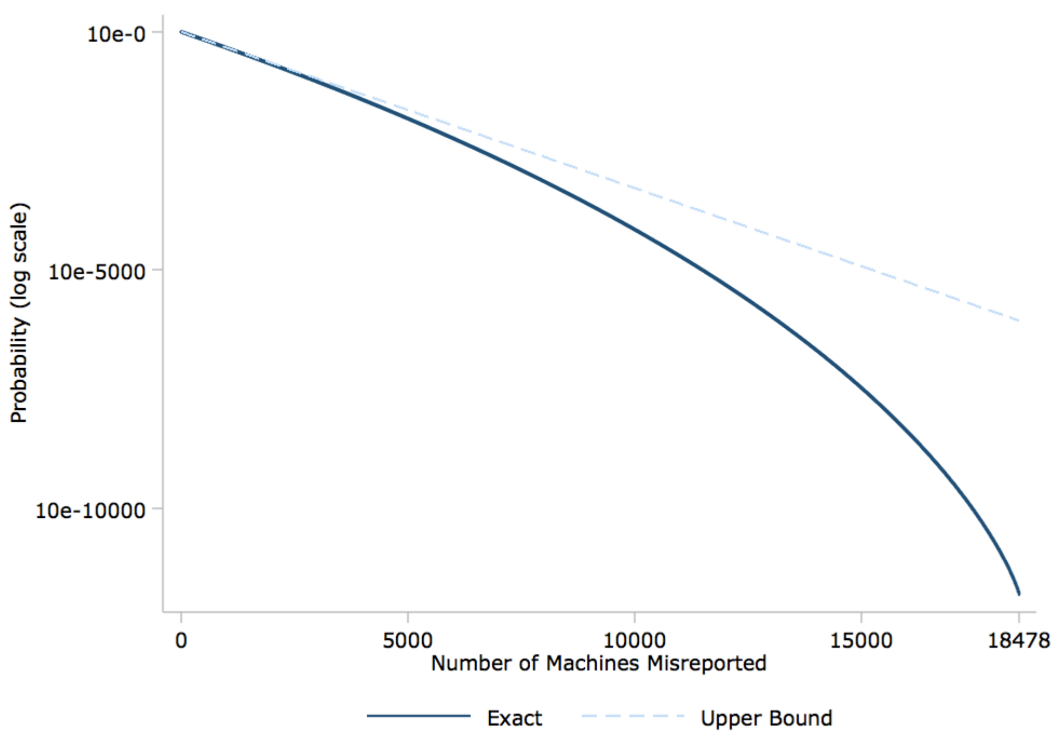
In other words, the probability of a completely clean audit is less than the probability that a machine does not get audited raised to the power of the number of errors. Obviously, a full audit $(n = N)$ would imply $p(0; K) = 0$ if $K > 0$—that is, an audit of 100 percent of the machines cannot come back clean if there are any errors. Note however, that as K increases, this upper bound becomes increasingly loose (the exact probabilities become much, much smaller than the last calculation would suggest.)

By way of example, suppose there are at least 100 machines $(N = 100)$ and 50 machines in error $(K = 50)$. If we audit half of the total number $(n = 50)$, then the probability of a clean audit is less than one in 2^{50}, or about one in 10^{15}, or one in a thousand trillion.

If the audit rate is higher, then the probability of a clean audit is that much smaller. For the actual audit, $N = 39{,}303$ so if 50 machines are in error $(K = 50)$, then a clean audit with $n = 20{,}825$ $(\frac{n}{N} \approx 0.53)$ has a probability of less than one in about 25,000 trillion.

Figure A1 shows how the probability of the observed clean audit varies with the number of possible machine errors in the population (K). Both the exact figures and the upper bound are shown. Note that the upper bound is only a good estimate of the probability for $K \lesssim 2500$, but the probabilities are already meaninglessly tiny for $K = 2500$—less than 1 in 10^{1886}.

Figure A1
Probability of a Clean Audit by Number of Machines Misreported



Source: Author’s calculations

References

Center for Economic and Policy Research. 2006. "Discrepancies and Lack of Transparency Mar Mexican Election Results." Center for Economic and Policy Research. Washington D.C. Available at: http://www.cepr.net/index.php/press-releases/press-releases/discrepancies-and-lack-of-transparency-mar-mexican-election-results/

Center for Economic and Policy Research. 2006a. "CEPR Adds up Available Recount Data, Finds Significant Vote Reduction for Calderon." Center for Economic and Policy Research. Washington D.C. Available at: http://www.cepr.net/index.php/press-releases/press-releases/cepr-adds-up-available-recount-data-finds-significant-vote-reduction-for-calderon/

Weisbrot, Mark, Luis Sandoval, and Carla Paredes-Drouet. 2006. "An Analysis of Discrepancies in the Mexican Presidential Election Results." Center for Economic and Policy Research. Washington D.C. Available at: http://www.cepr.net/index.php/publications/reports/an-analysis-of-discrepancies-in-the-mexican-presidential-election-results