13 FINDINGS OF AN INDEPENDENT PANEL ON ALLEGATIONS OF STATISTICAL EVIDENCE FOR FRAUD DURING THE 2004 VENEZUELAN PRESIDENTIAL RECALL REFERENDUM

PANEL PARTICIPANTS

Henry Brady

Class of 1941 Monroe Deutsch Professor of Political Science and Public Policy at the University of California, Berkeley, with appointments in the Department of Political Science and the Goldman School of Public Policy. Faculty director of Berkeley's Survey Research Center and the University of California Data Archive and Technical Assistance (UC DATA) program.

Dr. Brady has written extensively on research methods and statistical techniques for analyzing social science data as well as on social welfare policy, political participation, and voting systems.

Richard Fowles

Associate Professor of Economics, University of Utah.

Dr. Fowles is an econometrician specializing in model specification fragility and analysis and also works on low probability/high consequence decision-making and public policy economics.

Aviel Rubin

Professor of Computer Science and Technical Director of the Information Security Institute at Johns Hopkins University.

Dr. Rubin is the author of several books on computer security and co-author of a report that showed security flaws in a widely used electronic voting system and that focused a national spotlight on the issue.

Jonathan Taylor

Assistant Professor, Department of Statistics, Stanford University.

Dr. Taylor's work focuses on multiple comparisons problems, specifically in neuroimaging and smooth stochastic processes.

Nicholas Beauchamp (Panel Coordinator)

Carter Center consultant on the Venezuela project.

(Apart from Mr. Beauchamp's work at The Carter Center, none of the panel members had any connection to Venezuela or the recall referendum before Aug. 15. Some of them were contacted or consulted by Venezuelans or The Carter Center after the event.)

INTRODUCTION

Immediately following the Aug. 15, 2004, presidential recall referendum in Venezuela, a number of allegations of fraud began to appear in the Venezuelan press. Some of these allegations were based solely on statistical studies of the recall referendum returns, while other allegations examined that data in combination with exit polls, registered voters lists, the 2000 election results, or the results of the November 2003 petition to hold a recall referendum. Because the Aug. 18 audit of a sample of recall referendum returns found no significant discrepancy between the official electronic returns and the paper receipts printed out after each vote, some of the allegations of fraud extended to the audit as well. The Carter Center convened a panel of independent experts who had not been involved in the Venezuelan recall referendum to explore some of these statistically based allegations of fraud, and to make recommendations for minimizing the potential for fraud suspicions in future elections.

Because there have been numerous reports alleging evidence of fraud and because many of these reports have not yet been published or publicly released in full, the panel has only investigated a subset of the claims that have been made. The claims of fraud investigated fall into four categories. In general, the various allegations of fraud have not been integrated into a single scenario, and the evidence for each accusation has been treated on its own terms.

CLAIMS INVESTIGATED

1. Anomalous distributions of recall referendum votes among voting machines, including anomalously high rates of matching Yes votes among machines at a single voting table or voting center.

2. Unusual correlations between recall referendum results, exit polls, and the November 2003 petition to hold the recall referendum.

3. Regression analysis incorporating various factors such as Yes votes, 2003 petition signatures, and registered voters, showing that the audited voting centers behave slightly differently from the total universe of voting centers.

4. The failure of recall referendum vote totals to conform to "Benford's Law" governing the frequency of the first and second digits in those totals.

SUMMARY OF CONCLUSIONS

The Carter Center panel has found that none of the statistical studies examined here present evidence that fraud occurred during the 2004 presidential recall referendum:

1. The number of voting centers with matching machines is only slightly outside the expectations generated by most election models. A broader examination of vote distributions using more powerful tools has found no significant anomalies in the recall returns.

2. The panel rejects the hypothesis that a direct correlation between 1) the differences between the recall referendum and November 2003 petition results and 2) the differences between recall referendum and exit poll results implies that the referendum results were fraudulently manipulated. Instead, as others have suggested, there are numerous reasons why the petition and exit poll errors could be directly correlated, most of which have not been tested.

3. The panel has attempted to replicate the results of the Hausmann and Rigobon report claiming that the audited voting centers do not behave exactly the same in a linear regression model as the total universe of votes and that, therefore, the audited sample was not random. The panel found that this result was very dependent on the 2003 recall petition data and that slightly different petition data sets reduced or eliminated the audit anomaly. Because the anomaly is small and not robust, the panel concludes that these regressions do not present evidence that the audit sample was fraudulently biased.

4. The panel concludes that there is insufficient evidence that Benford's Law applies to election results in general. Furthermore, a simple but plausible model of the election does not produce results that conform to Benford's Law.

RECOMMENDATIONS FOR FUTURE ELECTIONS

To strengthen trust in electronic voting, there should be no doubt about the security of any stage of the process. The code running on voting machines should be publicly available for inspection as should the code running on any server involved in aggregating and tallying the votes. Vote-aggregating servers should neither be in contact with the Internet nor accessible via modem, except for the period when voting machines send in their tallies. As much as possible, the actions of these servers should be equally open to observation by any of the parties during the tallying phase. There should be no mechanism for any voting machine to receive instructions or modifications from the server, and only poll workers at the voting machines should be able to initiate the transfer of vote tallies to the server. The vote tally from each machine should also be recorded to a memory stick or other portable memory device, which should be physically delivered to the central electoral agency and compared with the electronically reported tallies. There should be no means for election workers to modify the results of voting machines, apart from resetting the machines at the beginning of the election day.

The paper voting receipts collected from each voter during the election are an excellent method for building trust in electronic voting by providing a verifiable paper record of the election. During the Aug. 18 audit, the paper record allowed a relatively easy comparison of the paper receipts with the official electronic election results. However, the three days between the election and the audit raised doubts about the authenticity of the paper voting records, doubts which could be largely eliminated by keeping the paper records under observation by both sides during the entire time between the election and any audit. In the future, there should be a clearly established chain of custody for the paper ballot receipts, one that allows the maximum degree of transparency to observation without jeopardizing the security of the paper records.

Doubts about the electronic tally and the paper receipts could further be diminished by shortening the time between the election and the audit, with the ideal scenario being a "hot" audit at a random sample of voting centers immediately after voting has closed. The sampled centers should be selected immediately after voting has ended in order not to affect voting or warn away fraud attempts at those machines. Such a hot audit was attempted during the Aug. 15 referendum but was not successful, largely due to he irregular closing times of voting centers. During the Oct. 31 election, however, a hot audit was successfully performed following most of these guidelines, and a sample of approximately 5,000 machines (one for each automated voting center) found a very close match between electronic and paper results. Because the audit was performed without delay and because various other security measures were implemented for the election (such as printing machine tallies before transmitting results), the potential for doubt was minimized, and few accusations of fraud later appeared. The Oct. 31 election and audit have set a good standard for establishing the security of electronic voting results.

As trust builds in the electronic voting system, it may become less necessary to conduct hot audits for every election. In those cases where all parties have agreed to forego a hot audit, it will be crucial to keep the paper voting records in well-observed locations until the period in which doubts could trigger a cold audit has passed. If ballot boxes and voting machines have not been under continuous observation since the election, any cold audit should send observers to all ballot box and voting machine storage locations prior to the selection of audited machines to prevent the audited receipts or machines from being manipulated between the announcement and the pickup. This technique, employed in the Aug. 18 audit, ensures that any attempt to manipulate the paper receipts to match fraudulent electronic tallies would require replacing the paper for every single ballot box, a job that would be quite difficult logistically as well as quite difficult to keep secret.

To remove doubts about the audited sample being truly random, the program code for generating the sample should be open for all to see; the computer on which the code is run should belong to a neutral party and may be reformatted prior to running the program; the numerical "seed" for the random sample generator should be chosen in pieces by each participating party, and these pieces should be combined by bitwise XOR rather than merely conjoining the pieces (bitwise XOR diminishes the possibility that any party's seed-piece might deliberately shape the outcome); for a cold audit, the sample-generating program should be run publicly, and its outcome publicly reported, while for a same-day hot audit, the parties may not wish to publicly identify the audited voting centers until after the election is over, in order not to affect the voting. Most of these guidelines were followed during the Aug. 18 audit, although only one party (the CNE) selected the seed, and the sample-generating program was run on a CNE computer (see the discussion for Claim 3 for more on this). These oversights can be easily corrected in future elections, as they were, to a large degree, in the Oct. 31 election.

Discussion

Claim 1

The first concrete allegations that there was statistical evidence of fraud were based on an examination of the vote totals for machines at the same voting table or voting center. Initially, the claim was that, because many machines at a given table appeared to have matching Yes vote totals, the Yes totals might have been capped by a hidden program in the voting machines which then transferred any votes above that cap to the No option. However, this specific scenario was quickly abandoned when it was shown that the matching totals were normally distributed and that the totals for the third machine at a table where the other two machines were "capped" had more Yes votes as often as it had fewer. This "capping" scenario was largely dropped and, instead, somewhat more rigorous claims were put forward that the frequency of matching machines was anomalously high, without committing to a specific fraud scenario.

The panel examined two public reports supporting these claims, by Valladares and by Jimenez et. al. The panel also looked more generally at the match between actual vote distributions and plausible models of the election. Regarding matching totals, the panel largely agrees with the results in Felten et. al., Hausmann and Rigobon, and Taylor, finding that matching totals were in some cases slightly higher than predicted by some models but that usually the matching rate was within the expected range generated by reasonable election models.

To further examine this issue, a closer examination of the multinomial deviance under the Poisson model was conducted by Taylor using a "False Discovery Rate" (FDR) analysis to see if the minor departures from the Poisson model were due to manipulation of some nontrivial subset of machines. This analysis presupposes the presence of an unusual dispersion in the Yes results that could manifest itself as an unusual number of ties or other types of vote manipulation and then tries to estimate the departure from a global null model. In this case, the null model was one of the models proposed by Valladares that assumes that the votes within a table are independent Poisson random variables with equal parameters across each machine. (While the final model of Valladares et al. assigned the parameters slightly differently, the panel feels this would not greatly affect the FDR analysis.) Taylor concludes that those small discrepancies that can be found are due mainly to 21 voting tables in one analysis and due to eight tables in another. In general, the dispersions of the No and Yes votes seem to be the same, and he concludes that the data show "no clear departures from the Poisson model."

Claim 2

Part 1 of Hausmann and Rigobon's In Search of the Black Swan concerns the recall referendum results in general, while Part 2 deals with the Aug. 18 audit. Part 1 will be discussed in this section (Claim 2), while Part 2 will be discussed in the following section (Claim 3).

In Part 1 of their report, Hausmann and Rigobon use linear regression models of Yes votes versus exit polls and Yes votes versus signatures collected during the November 2003 petition to hold a referendum and compare the residuals for those regressions, finding that they are correlated at 0.17 ($p \sim 0.1\%$). They assume that these residuals should not be directly correlated (because the exit polls and signature collection took place under different circumstances) and conclude that therefore any correlation is due to a hidden fraud factor within the residual terms arising from manipulation of the Yes vote. However, even accepting this correlation, there are numerous explanations for why the exit poll and signature residuals could be directly correlated. For instance, two voting centers may have had the same percentage of opposition voters, but at one center, those voters may have been more motivated and dedicated to the cause, industriously signing the more complex petitions at higher rates and later seeking out exit pollsters to produce a higher Yes exit poll response in those areas. Similarly, since the petition and the exit polls in question were staffed by the same local opposition supporters, areas with more industrious opposition workers could have gathered more signatures and later (presuming a polling bias toward opposition respondents, as the exit polls suggest) gathered more Yes exit poll responses. Without testing these and related hypotheses, the correlation between exit poll and signature residuals cannot be taken as evidence of fraud.

Claim 3

In Part 2 of their report, Hausmann and Rigobon present evidence that audited parroquias (voting precincts) in the 2004 Venezuelan referendum are statistically different from the unaudited ones. Their claim is that these results cast doubt on the reliability of the audit and show that the audited sample was not truly random and representative of the entire universe of voting centers. However, in attempting to replicate Hausmann and Rigobon's results, Richard Fowles found that the data used to assess fraud is fragile – it does not clearly show whether there is a difference between the audited and unaudited precincts.

Hausmann and Rigobon's ordinary least squares (OLS) regression model demonstrates that in both audited and unaudited precincts there is a positive relationship between the number of signatures obtained and the number of Yes votes. They claim, however, that the presence of fraud introduces a bias in OLS estimation that will unambiguously result in a positive and statistically significant coefficient on a particular variable defined in their regression model. That explanatory variable is created using a binary variable indicating whether or not a precinct was audited, multiplied by the logarithm of the number of signatures collected during the referendum. In this model, the dependent variable is the logarithm of the number of Yes votes in the precinct. This model directly estimates the elasticity of the number of Yes votes with respect to the number of petition signatures. In Table 10 of the Hausmann and Rigobon paper, this elasticity is estimated at .958 for unaudited precincts and .105 higher than this for audited precincts.

During attempts to replicate the Hausmann and Rigobon results, two basic data sets were used. One was supplied by Rigobon to Richard Fowles, and the other came from the civil society group Súmate (via The Carter Center). Although the numbers in these data sets were close, they were not exactly the same in the key variables needed to replicate the Hausmann and Rigobon results. Differences were present for the numbers of signatures collected in areas corresponding to each voting center (see below), and slight differences were present in the data specifying which centers had been audited. The Carter Center also supplied its own data sets for signatures and the audit status of a parroquia. Thus there were nine different combinations of the key explanatory variable and two combinations of Yes votes that could be used to assess the reliability of the Hausmann and Rigobon result. Our regressions revealed that the statistical significance of the estimated coefficients highlighted by Hausmann and Rigobon was due to differences in the signature data from the two data sets.

For example, a pure Rigobon regression used the Rigobon Yes vote, the Rigobon binary audit variable, and the Rigobon signature data. As expected, this regression results in a statistically significant estimated coefficient on the key variable of interest. A pure Súmate regression (using Súmate's Yes, audit, and signature data) does not result in an estimated coefficient on the key variable that is statistically different from zero and thus presents no evidence of fraud under the Hausmann and Rigobon assumptions. A mixed regression used assorted combinations of the Súmate and Rigobon data. If, for example, the Súmate signature variable was used in an otherwise pure Rigobon regression, the Hausmann and Rigobon results did not show up. Similarly, if the Rigobon signature variable was used in an otherwise pure Súmate regression, the Hausmann and Rigobon result appeared.

The differences in the signature data may be due to the difficulty of assigning signers in 2003 to voting centers in 2004, since the voting locations were shifted somewhat in between these two dates, requiring a careful reallocation of signatures to the new voting centers. But in any case, the panel concludes that the data used to assess this fraud is fragile and is not conclusive evidence of fraud in the audit. This conclusion is independent of any criticism of the Hausmann and Rigobon assumptions that they use to focus attention on their regression, and it is also independent of any assessment of basic fragility in model specification.

Further Discussion of the Aug. 18 Audit

Given that the audited paper voting receipts matched the electronic tallies for those machines extremely closely, most fraud scenarios must include the hypothesis that the audit was, in fact, not random and that the CNE was able to force the selection of the audited sample. This could have been done by carefully selecting the seed for the program generating the random sample. However, the CNE did not expect to select the entire seed itself – this contingency was only made possible when the opposition refused to participate in the seed selection, the OAS and Carter Center observers subsequently declined to choose the seed, and the CNE officials therefore had to select the seed themselves. The CNE would also have needed to run many seeds through the program ahead of time in order to produce a sample whose overall results matched a fraudulent No victory.

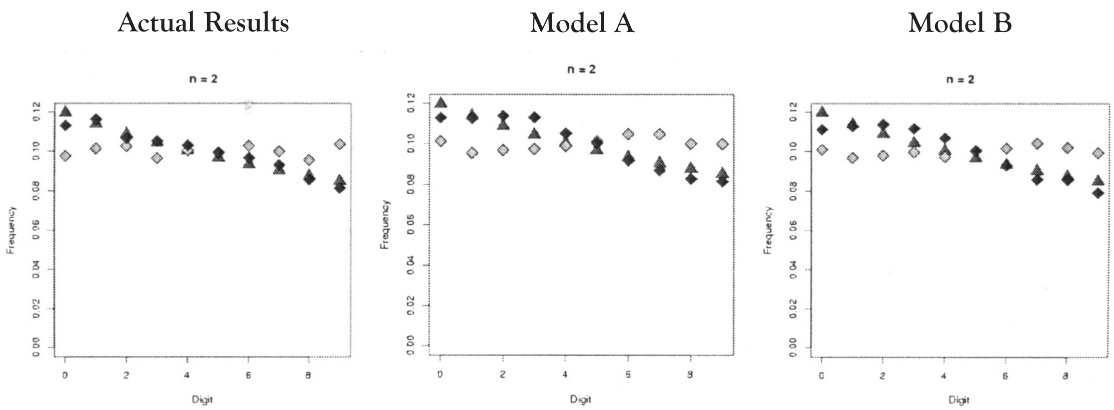
Alternatively, some have claimed that the program was simply hacked. The Carter Center and OAS checked the program before and after the drawing of the sample and found that it produced the same sample given the same seed. The Carter Center further checked the program with many different seeds and found that every machine in the universe did appear in at least one sample drawn, indicating that the sample was not restricted to a subset of the machines. While there can be no absolute guarantee that code run on an insecure computer may not have been briefly modified or that the seed may have been chosen to produce a specific set of centers, there is at least one additional reason to believe that this did not occur. On the hypothesis that Yes actually won, a sample of voting centers that supported a resounding No victory would necessarily be made up of anomalously No-leaning areas. To examine this possibility, Weisbrot et. al. looked at whether the audited centers were anomalously pro-Chávez in the 2000 election. Instead, they found that the audited centers were generally representative in the 2000 election. It remains conceivable that seeds could have been tried until a sample was chosen that was both representative in 2000 and unusually pro-Chávez in 2004 – but the panel finds this unlikely.

As mentioned above, to prevent future concerns regarding the randomness of the audit, any audit should begin immediately after voting has closed; the program code should be open for all to examine; the program should run on a secure, neutral computer; and the seed should be publicly chosen by a variety of parties who combine their pieces in an XOR fashion. Additionally, although some of the security measures recommended here have yet to be implemented, the Oct. 31 hot audit successfully eliminated most of the potential doubts about the electronic ballot tallies.

Claim 4

More recently, claims have been made by Mikoss and by Pericchi and Torres that a comparison of the recall referendum results with "Benford's Law" shows that the results are fraudulent. Benford's Law is occasionally used to search for financial fraud and governs the rate that the digits 0-9 appear in some sets of data. Pericchi and Torres argue that the election data (specifically, the second digit of machine-level totals) conform to Benford's Law, except for the automated No results. Mikoss examines the No minus Yes totals and finds that it matches Benford's Law but that when one switches a varying percentage of votes from No "back" to the hypothetical Yes victory, the match with Benford's Law is much improved at a 24 percent switch.

The panel believes that there are many reasons to doubt the applicability of Benford's Law to election returns. In particular, Benford's Law is characteristic for scale-invariant data, while election machines are allocated to maintain a relatively constant number of voters per machine. Brady finds, for example, that the first digit of precinct-level electoral data for Cook County, the city of Chicago, and Broward County, Fla., depart significantly from Benford's Law, primarily because of the relatively constant number of voters in voting precincts. He also describes a set of assumptions that fit many kinds of electoral data (at least approximately) and that will lead to distributions other than Benford's. Brady also finds that taking the difference



Graphs are of Digits versus Frequency. Dark diamond=Benford's Law, Light diamond=Yes, Triangle=No

of votes for two major presidential candidates in Cook County precincts leads to an empirical distribution that departs from Benford's Law (which Mikoss takes in the analogous case as evidence of fraud). Furthermore, Brady finds that in situations where some of the conditions for the application of Benford's Law seem to apply, the electoral data fail to conform to the law. In short, Benford's Law does not generally apply to electoral data and even in cases where we suspect that it might apply, we find that it does not. All in all, Benford's Law seems like a very weak instrument for detecting voting fraud. There are many reasons to believe that it does not apply to electoral data, and empirical tests suggest that deviations from the law are not necessarily indicative of fraud.

It also seems reasonable to assume that if Benford's Law does hold for election data, then it should hold for a reasonable model of the election. Jonathan Taylor tested Benford's Law for two models of the election: a) a simple Poisson model with parameter varying by table, with Yes/No votes split 40 percent Yes/60 percent No across all mesas and b) Poisson with a parameter varying by table, and Yes/No votes split equally among machines, which is the same model used for the FDR analysis above and was one of the models proposed by Valladares et al. as a model of the election data without fraud. The results for the second digit are as follows:

Clearly the digit frequencies for both models match the actual data fairly closely, and the "violation" of Benford's Law for the second digit of No votes does not show fraud, particularly if one accepts the claim that Valladares et al.'s model is a model of a fair election without fraud.

CONCLUSIONS

The panel finds that none of the reports examined present evidence that there was significant fraud during the Aug. 15 presidential recall referendum.

Though the rate of Yes vote matches at times is somewhat greater than that predicted by some models of the election, a thorough examination of vote distributions finds no significant difference between a reasonable model of the election and the actual results.

Hausmann and Rigobon's comparison of exit polls, petition signatures, and recall referendum results finds a direct correlation between the first two, which they attribute to fraud but which the panel considers to be a simple correlation potentially due to any number of nonfraudulent causes.

A regression analysis showing that the audited sample behaves slightly differently than the total universe of votes proves to be dependent on the 2003 petition signature data, and that result was not robust under small changes in that data.

Alleged evidence of fraud based on Benford's Law instead demonstrates how closely the election data match reasonable models of the election.

Although doubts were raised about the security of ballot boxes in between the recall referendum and the audit, and doubts were raised about the security of the program generating the audit sample, it should be emphasized that the paper audit produced results that very closely matched the electronic tally and that any manipulation of the non-audited centers would not only need to bias the audited centers to match a fraudulent outcome but would need to choose centers that were somehow also representative in the 2000 election. It might also be mentioned that none of the claims for evidence of fraud suggested a fraud so great as to change the exit-polled 60/40 opposition win to the official 40/60 government win.

That said, the panel only concludes that there is no statistical evidence of fraud based on the reports we have examined. The panel cannot explain why the exit polls proved so mistaken– though, following Weisbrot, the panel can point to one exit poll conducted without opposition help by the American firm Evans/McDonough that found results of 55 percent No to 45 percent Yes.

The Venezuelan recall referendum implemented more security and trust-building features than electronic elections in many other countries, including paper receipts and the cold audit. Although the hot audit on Aug. 15 was not successful, the hot audit performed during the Oct. 31 elections was quite successful, dispelling many of the kinds of doubts that appeared after the recall referendum. The panel has made further recommendations in the section above for building trusted elections, but the Venezuelan election authority already has most of the pieces in place for building a trustworthy voting system in which it will be even more difficult to perpetrate any substantial fraud.

WORKS CITED

Brady, Henry E. *Comments on Benford's Law and the Venezuelan Election*, January 19, 2005. (University of California, Berkeley)

Felten, Edward W., Aviel D. Rubin and Adam Stubblefield, *Analysis of Voting Data from the Recent Venezuela Recall Referendum*, September 1, 2004. (Johns Hopkins University and Princeton University)

Jiménez, J., R. Jiménez and A. Marcano. *¿Cuan probables son las coincidencias observadas en el referendum revocatorio Venezolano?* September 9, 2004. (Universidad Central de Venezuela and Universidad Simón Bolívar)

Hausmann, Ricardo and Roberto Rigobon, *En busca del cisne negro: Analisis de la evidencia estadistica sobre fraude electoral en Venezuela*, September 3, 2004. (Harvard University and MIT)

Mikoss, M.Sc. Imre. *Evidencia de Manipulación Artificial de los Resultados al Aplicar la Ley de Benford al Referedum Venezolano de Agosto de 2004*. (Universidad Simón Bolívar, Valle de Sartenjas, Departamento de Física, Caracas, Venezuela)

Pericchi, Luis Raúl and David Torres. *La Ley de Newcomb- Benford y sus aplicaciones al Referendum Revocatorio en Venezuela. Reporte Técnico no- definitivo 2a. version: Octubre 1, 2004*. (Universidad de Puerto Rico and Universidad Simón Bolívar)

Taylor, Jonathan, *Too Many Ties? An empirical analysis of the Venezuelan recall referendum counts*, September 10, 2004 (Stanford University)

Valladares, Elio, *Segundo avance del informe sobre mesas donde dos o mas de sus cuadernos presentan el mismo numbero de votos por la opcion "SI", September 2004* (University of Virginia)

Weisbrot, Mark, David Rosnick and Todd Tucker, *Black Swans, Conspiracy Theories, and the Quixotic Search for Fraud: A look at Hausmann and Rigobon's Analysis of Venezuela's Referendum Vote, September 20, 2004*. (Center for Economic and Policy Research)