Detecting Voter Fraud in an Electronic Voting Context

An Analysis of the Unlimited Reelection Vote in Venezuela

Ines Levin  Gabe Cohn
Peter Ordeshook  Michael Alvarez

California Institute of Technology

August 10, 2009
1. Introduction: Relevance of Election Forensics
2. Previous Research
3. Electronic Voting System in Venezuela
4. The 2007 and 2009 Constitutional Referendums
5. Data, Methodology and Results
6. Conclusion
Election Integrity and Electronic Voting

- Election integrity is important for the elected government to have legitimacy, and to prevent political conflict.
Election Integrity and Electronic Voting

- Election integrity is important for the elected government to have legitimacy, and to prevent political conflict.
- There are concerns about election integrity as new technologies are introduced.
Election integrity is important for the elected government to have legitimacy, and to prevent political conflict.

There are concerns about election integrity as new technologies are introduced.

Specifically, voting machines are sometimes perceived as black boxes, and may be vulnerable to hacker attacks, to erroneous or malicious software, or to manipulation by vendors with partisan political interests (Kohono et al. 2004; State of Alaska Division of Elections 2007, 2008; California Secretary of State 2007; Ohio Secretary of State 2007).
Election integrity is important for the elected government to have legitimacy, and to prevent political conflict.

There are concerns about election integrity as new technologies are introduced.

Specifically, voting machines are sometimes perceived as black boxes, and may be vulnerable to hacker attacks, to erroneous or malicious software, or to manipulation by vendors with partisan political interests (Kohono et al. 2004; State of Alaska Division of Elections 2007, 2008; California Secretary of State 2007; Ohio Secretary of State 2007).

Pre- and post-election audits, as well as election forensic tools are important because they compensate for the black box nature of electronic voting.
Most of the forensic tools applied in our paper were previously developed to study election fraud in Russia and Ukraine (see Myagkov, Ordeshook and Shakin 2009).
Aim of the Study and Previous Research

- Most of the forensic tools applied in our paper were previously developed to study election fraud in Russia and Ukraine (see Myagkov, Ordeshook and Shakin 2009).

- These tools are designed to detect artificial election heterogeneity, unexplained by socio-demographic factors.
Aim of the Study and Previous Research

- Most of the forensic tools applied in our paper were previously developed to study election fraud in Russia and Ukraine (see Myagkov, Ordeshook and Shakin 2009).
- These tools are designed to detect artificial election heterogeneity, unexplained by socio-demographic factors.
- Ukraine 2004 offered the perfect social science experiment when the presidential runoff was repeated twice within a month period, with the same candidates, issues and electorate, but fewer opportunities for electoral fraud.
Aim of the Study and Previous Research

- Most of the forensic tools applied in our paper were previously developed to study election fraud in Russia and Ukraine (see Myagkov, Ordeshook and Shakin 2009).

- These tools are designed to detect *artificial* election heterogeneity, unexplained by socio-demographic factors.

- Ukraine 2004 offered the perfect social science experiment when the presidential runoff was repeated twice within a month period, with the same candidates, issues and electorate, but fewer opportunities for electoral fraud.

- In this paper, we compare election returns in the 2007 and 2009 Venezuela referendums, where relatively similar constitutional reforms were put to a vote twice during a short period of time – approximately 1 year. Thus, similar to the Ukrainian case, the unlimited reelection vote in Venezuela offers an excellent opportunity for applying our set of forensic indicators.
Electronic Voting System in Venezuela

- Introduced before the 2004 presidential recall referendum.
Electronic Voting System in Venezuela

- Introduced before the 2004 presidential recall referendum.
- Has been used in 5 national elections – 3 constitutional referendums, 1 presidential election, and 1 parliamentary election.
Detecting Voter Fraud in an Electronic Voting Context

Introduction: Relevance of Election Forensics

Previous Research

Electronic Voting System in Venezuela

The 2007 and 2009 Constitutional Referendums

Data, Methodology and Results

Conclusion

Electronic Voting System in Venezuela

- Introduced before the 2004 presidential recall referendum.
- Has been used in 5 national elections – 3 constitutional referendums, 1 presidential election, and 1 parliamentary election.
- Security features (Carter Center 2007):
  - Encryption of voting information.
  - Randomization of information to prevent reconstruction of voting sequence.
  - Disabling unnecessary physical access points
  - Chain of custody procedures
Detecting Voter Fraud in an Electronic Voting Context

Introduction:
Relevance of Election Forensics
Previous Research
Electronic Voting System in Venezuela
The 2007 and 2009 Constitutional Referendums
Data, Methodology and Results
Conclusion

Touch Screen System (Smartmatic)

Source: BBC Mundo

2009 Referendum

Source: Sumate
2007 and 2009 Constitutional Referendums

- Some contents of the 2007 referendum
  - Removal of presidential reelection term limits.
  - Abolishing the autonomy of the central bank.
  - Expropriation of large land estates.

- The opposition alleged that some of these reform proposals were unconstitutional.

- The referendum failed to pass in 2007, but Chavez managed to obtain court approval for holding a new referendum in 2009.

- The 2009 referendum proposed the elimination of reelection term limits for public offices, including the presidency.

- The opposition argued that this was a mere repackaging of some of the 2007 proposals, and that it was unconstitutional – the same referendum proposal cannot be put to a vote more than once during the same National Assembly period.
The analysis of individual election figures was done using *mesa* (voting table) level data.
The analysis of individual election figures was done using \textit{mesa} (voting table) level data.

The comparison of election returns across elections was done using polling place level data, after exactly matching 8,815 polling places which were open during the three elections of interest.
Data

- The analysis of individual election figures was done using *mesa* (voting table) level data.
- The comparison of election returns across elections was done using polling place level data, after exactly matching 8,815 polling places which were open during the three elections of interest.
- The average number of voters per polling place in 2007 and 2009 was 1,230 and 1,744, respectively.
Data

- The analysis of individual election figures was done using *mesa* (voting table) level data.
- The comparison of election returns across elections was done using polling place level data, after exactly matching 8,815 polling places which were open during the three elections of interest.
- The average number of voters per polling place in 2007 and 2009 was 1,230 and 1,744, respectively.
- Source: In the case of the 2007 and 2009 referendums, we used a Python script to download the data from the web-page of the National Elections Council (CNE). Also, we downloaded 2006 election returns from the ESDATA web-page.
Detecting Voter Fraud in an Electronic Voting Context

Introduction: Relevance of Election Forensics

Previous Research

Electronic Voting System in Venezuela

The 2007 and 2009 Constitutional Referendums

Data, Methodology and Results

Conclusion

Election Returns: Ternary Plots

2006

2007

2009
Detection of anomalies in the distribution of turnout. If turnout varied based on purely random reasons, unrelated to socio-demographic factors or political attitudes, then we’d expect the distribution of turnout to look approximately normal. Thus, examination of the distribution of turnout in relatively homogeneous data can be used to detect artificial heterogeneity.
Results: Turnout Distribution

Figure: BLUE: 2006, RED: 2007, GREEN: 2009
Turnout Distribution: The Russian Case

Source: Myagkov, Ordeshook and Shakin (2009, Table 3.6)
Detection of anomalies in the distribution of the last two digits of precinct-level election returns. If election protocols were manipulated and filled in a non-random manner, unrelated to actual votes being cast, then we would expect heuristics for filling out protocols (see Berber and Scacco 2008, Shpilkin 2008), such as:

- Rounding-off of the last digit to 0 and 5.
Detection of anomalies in the distribution of the last two digits of precinct-level election returns. If election protocols were manipulated and filled in a non-random manner, unrelated to actual votes being cast, then we would expect heuristics for filling out protocols (see Berber and Scacco 2008, Shpilkin 2008), such as:

- Rounding-off of the last digit to 0 and 5.
- Avoiding rounding-off of the last digit to 0 and 5.
Forensic Tools (II): Distribution of Last Two Digits

- Detection of anomalies in the distribution of the last two digits of precinct-level election returns. If election protocols were manipulated and filled in a non-random manner, unrelated to actual votes being cast, then we would expect heuristics for filling out protocols (see Berber and Scacco 2008, Shpilkin 2008), such as:
  - Rounding-off of the last digit to 0 and 5.
  - Avoiding rounding-off of the last digit to 0 and 5.
  - Avoiding sequences of paired-numbers – There is experimental evidence that when people is asked to write sequences of random numbers, they tend to write down paired numbers less frequently (Chapanis 1953, Rath 1966, Boland and Hutchinson 2000).
Detecting Voter Fraud in an Electronic Voting Context

Introduction: Relevance of Election Forensics

Previous Research

Electronic Voting System in Venezuela

The 2007 and 2009 Constitutional Referendums

Data, Methodology and Results

Conclusion

Results: Distribution of Last Two Digits

<table>
<thead>
<tr>
<th>Last Digit 2006 (“Chavez” option)</th>
<th>Last Digit 2007 (“yes” option)</th>
<th>Last Digit 2009 (“yes option”)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Last 2 Digits 2006 (“Chavez” option)</th>
<th>Last 2 Digits 2007 (“yes” option)</th>
<th>Last 2 Digits 2009 (“yes option”)</th>
</tr>
</thead>
</table>
Forensic Tools (III): Regression Analysis of Vote Flows

- **Regression analysis of the flow of votes between elections** The idea is to estimate a contingency table showing the proportion of votes received by each alternative in the current election proceeding from each alternative in the previous election, including those who abstained.
Forensic Tools (III): Regression Analysis of Vote Flows

- *Regression analysis of the flow of votes between elections* The idea is to estimate a contingency table showing the proportion of votes received by each alternative in the current election proceeding from each alternative in the previous election, including those who abstained.

- **Model:**

  \[
  y_{1i} = \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}(1 - x_{1i} - x_{2i})
  \]

  \[
  y_{2i} = \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}(1 - x_{1i} - x_{2i})
  \]

  \[
  y_{3i} = \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}(1 - x_{1i} - x_{2i})
  \]
Forensic Tools (III): Regression Analysis of Vote Flows

- **Regression analysis of the flow of votes between elections** The idea is to estimate a contingency table showing the proportion of votes received by each alternative in the current election proceeding from each alternative in the previous election, including those who abstained.

- **Model:**
  
  \[
  y_{1i} = \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}(1 - x_{1i} - x_{2i}) \\
  y_{2i} = \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}(1 - x_{1i} - x_{2i}) \\
  y_{3i} = \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}(1 - x_{1i} - x_{2i})
  \]

- Coefficients should lie between 0 and 1, but this may not be the case due to aggregation bias (King 1997) or electoral fraud (Myagkov, Ordeshook and Shakin 2009).
Forensic Tools (III): Regression Analysis of Vote Flows

- **Regression analysis of the flow of votes between elections**
  The idea is to estimate a contingency table showing the proportion of votes received by each alternative in the current election proceeding from each alternative in the previous election, including those who abstained.

- **Model:**

  \[
  y_{1i} = \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}(1 - x_{1i} - x_{2i}) \\
  y_{2i} = \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}(1 - x_{1i} - x_{2i}) \\
  y_{3i} = \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}(1 - x_{1i} - x_{2i})
  \]

- Coefficients should lie between 0 and 1, but this may not be the case due to aggregation bias (King 1997) or electoral fraud (Myagkov, Ordeshook and Shakin 2009).

- In this paper we estimate a model with random coefficients by region, which should help mitigate aggregation issues. We do not restrict the coefficients because this would prevent us from detecting unreasonable figures caused by electoral fraud.
Results: Regression Analysis of Vote Flows

Flow of votes between 2007 and 2009

<table>
<thead>
<tr>
<th></th>
<th>Yes 07</th>
<th>No 07</th>
<th>Abst 07</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes 09</td>
<td>0.97</td>
<td>-0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>No 09</td>
<td>-0.11</td>
<td>0.94</td>
<td>0.14</td>
</tr>
<tr>
<td>Abst 09</td>
<td>0.13</td>
<td>0.10</td>
<td>0.56</td>
</tr>
<tr>
<td>Vertical sum</td>
<td>1.00</td>
<td>0.96</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Main observations:

- Voters who supported the yes option in 2007, overwhelmingly support the yes option in 2009.
- Voters who supported the no option in 2007, overwhelmingly support the no option in 2009.
- New turnout is split 2 to 1 between the yes and no alternatives in 2009.
Results: Regression Analysis of Vote Flows

<table>
<thead>
<tr>
<th>Flow of votes between 2007 and 2009</th>
<th>Yes 07</th>
<th>No 07</th>
<th>Abst 07</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes 09</td>
<td>0.97</td>
<td>-0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>No 09</td>
<td>-0.11</td>
<td>0.94</td>
<td>0.14</td>
</tr>
<tr>
<td>Abst 09</td>
<td>0.13</td>
<td>0.10</td>
<td>0.56</td>
</tr>
<tr>
<td>Vertical sum</td>
<td>1.00</td>
<td>0.96</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Main observations:
- Voters who supported the yes option in 2007, overwhelmingly support the yes option in 2009.
- Voters who supported the no option in 2007, overwhelmingly support the no option in 2009.
Results: Regression Analysis of Vote Flows

<table>
<thead>
<tr>
<th>Flow of votes between 2007 and 2009</th>
<th>Yes 07</th>
<th>No 07</th>
<th>Abst 07</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes 09</td>
<td>0.97</td>
<td>-0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>No 09</td>
<td>-0.11</td>
<td>0.94</td>
<td>0.14</td>
</tr>
<tr>
<td>Abst 09</td>
<td>0.13</td>
<td>0.10</td>
<td>0.56</td>
</tr>
<tr>
<td>Vertical sum</td>
<td>1.00</td>
<td>0.96</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Main observations:

- Voters who supported the yes option in 2007, overwhelmingly support the yes option in 2009.
- Voters who supported the no option in 2007, overwhelmingly support the no option in 2009.
- New turnout is split 2 to 1 between the yes and no alternatives in 2009.
Forensic Tools (IV): Predicted Vote vs. Actual Vote

- Assuming voter preferences are *in equilibrium*, we can estimate a model of voter choice in a previous election, and then use it to predict voter choice in the current election. Then, we can determine which precincts deviate significantly from expected values, and use that as a signal of election irregularities (Alvarez and Katz 2008).
Results: Predicted Vote vs. Actual Vote in 2009
Forensic Tools (V): Turnout vs. Vote Share of Eligible Electorate

- Detection of anomalies in the relationship between turnout and vote shares. If turnout were unrelated to a candidate's support, then the slope of a univariate regression between a candidate’s share of the eligible electorate and turnout should approximate the candidate’s overall proportion of the total turnout – i.e., it shouldn’t be larger than 1 or smaller than 0.
Results: Turnout vs. Vote Share of Eligible Electorate

**“Yes” Victory in 2007**

**“No” Victory in 2007**

**Turnout 2009** vs. **Votes / Eligible Electorate 2009**

- **BOLIVAR**
- **CARABOBO**
- **COJEDES**
- **DELTA AMACURO**

The graph shows the relationship between turnout and vote share for eligible electors in 2009, categorized by the victory in 2007.
Conclusion

- Studying the integrity of electronic elections is a challenging task – manipulation of the software code can lead to types of fraud difficult to detect with some of our forensic indicators.
Conclusion

- Studying the integrity of electronic elections is a challenging task – manipulation of the software code can lead to types of fraud difficult to detect with some of our forensic indicators.

- Elections taking place within a short period of time, with similar electorates and issues, offer an excellent opportunity for applying tools designed for detecting anomalies and outliers in election returns.

We did not find evidence of widespread electoral fraud in the last two Venezuela referendums – such as, generalized support for the 'yes' alternative in 2009 in polling places where the 'no' alternative won in 2007 – but we did detect anomalies and outliers in different stages of our analysis.

Results should be interpreted by experts on the electoral context of the different regions, and used as a complement to pre- and post-election audits, as well as reports from election observers.
Conclusion

- Studying the integrity of electronic elections is a challenging task – manipulation of the software code can lead to types of fraud difficult to detect with some of our forensic indicators.

- Elections taking place within a short period of time, with similar electorates and issues, offer an excellent opportunity for applying tools designed for detecting anomalies and outliers in election returns.

- We did not find evidence of widespread electoral fraud in the last two Venezuela referendums – such as, generalized support for the ‘yes’ alternative in 2009 in polling places where the ‘no’ alternative won in 2007 – but we did detect anomalies and outliers in different stages of our analysis.
Conclusion

- Studying the integrity of electronic elections is a challenging task – manipulation of the software code can lead to types of fraud difficult to detect with some of our forensic indicators.

- Elections taking place within a short period of time, with similar electorates and issues, offer an excellent opportunity for applying tools designed for detecting anomalies and outliers in election returns.

- We did not find evidence of widespread electoral fraud in the last two Venezuela referendums – such as, generalized support for the ‘yes’ alternative in 2009 in polling places where the ‘no’ alternative won in 2007 – but we did detect anomalies and outliers in different stages of our analysis.

- These anomalies are not, by themselves, proof of electoral fraud. Results should be interpreted by experts on the electoral context of the different regions, and used as a complement to pre- and post- election audits, as well as reports from election observers.
Future Research

- Develop tests of hypothesis associated with some of the forensic tools, to determine to what extent deviations from our expectations constitute strong signals of electoral fraud.
Future Research

- Develop tests of hypothesis associated with some of the forensic tools, to determine to what extent deviations from our expectations constitute strong signals of electoral fraud.

- Improve the regression analysis of the flow of votes. In particular, we’d like to:
  - Take care of the compositional nature of the dependent variables by employing the methodology suggested by Katz and King (1999).
Future Research

- Develop tests of hypothesis associated with some of the forensic tools, to determine to what extent deviations from our expectations constitute strong signals of electoral fraud.

- Improve the regression analysis of the flow of votes. In particular, we’d like to:
  - Take care of the compositional nature of the dependent variables by employing the methodology suggested by Katz and King (1999).
  - Address the ecological inference problem by specifying second-level models to explain heterogeneity of random coefficients across regions, taking socio-demographic factors and spatial-correlations into account (Gotway and Young 2004; Calvo and Escolar 2004; Haneuse and Wakefield 2004).