Evaluation of IREC's Statistical Analysis and Claims

Kenyans for Peace, Truth, and Justice

1 October 2008

Introduction

IREC's mandate suggested that an answer to the key question – who won the presidential elections – was forthcoming. "The terms of reference shall be to [...] (e) investigate the vote counting and tallying for the entire election with special attention to the presidential elections in order to assess the integrity of the results and make recommendations for improvements, adjustments or overhaul of the system." At the very least, Kenyans hoped that IREC would uncover whether and how intentional fraud was perpetrated, and, if possible, apportion blame for that fraud. In the minds of many, verifying the integrity of results could not be separated from identifying weaknesses in the organization and personnel of the ECK. Just as a doctor diagnoses a disease by observing the symptoms of a sick patient, IREC's mandate was to examine a large body of primary evidence in detail, looking for patterns suggestive of the organizational and personnel-related causes of the failed election process.

Unfortunately, the commission's approach failed to meet the challenge confronting them. From our perspective, their failure is rooted in two key aspects of their approach to the investigation. First, IREC either remained ignorant of or chose not to employ statistical techniques for forensic analysis of elections. Such methods are neither new, nor untested, having been applied in multiple electoral disputes in the past. Second, IREC did not think critically about why and how fraud could have been committed at KICC. As a result, IREC's "research design" for making a valid inference about the existence or non-existence of electoral fraud at KICC was insufficient.

This document outlines the specific mistakes, assumptions, and choices that IREC made in its report. Our goal is to point out where IREC went wrong, why, and how their report fails to address the questions raised in their mandate. IREC states: "'The integrity of the process and the credibility of the results were so gravely impaired by these manifold irregularities and defects that it is irrelevant whether or not there was actual rigging at the national tally centre. The results are irretrievably polluted."[1] We demonstrate several things below. First, IREC's investigation – specifically its statistical analysis – was insufficient to draw the conclusion that all of the ECK's results were wrong. Second, IREC failed to apply modern statistical techniques to the ECK's data, in effect excluding potentially important

[1] Executive Summary, p. x.

evidence from the consideration of the commissioners and the Kenyan people. Third, IREC's research design – specifically its decision to examine 19 problematic constituencies – was not capable of determining whether or not rigging occurred at KICC. Fourth, given its resources and time-frame, IREC could have conducted research capable of answering the questions at hand; they either remained ignorant of the proper methods for doing so, or deliberately chose not to pursue an effective analytical strategy. We hope that this analysis will provide Kenyans with enough information to decide for themselves whether or not to accept IREC's final report as accurate, or whether they have been short-changed by the process.

Sampling problems

Problem 1 ***IREC adamantly and repeatedly stated that statistical analysis should not be carried out on the electoral data: "[T]he IREC analysis of tallying etc. in nineteen sample constituencies demonstrates convincingly that there are so many more or less erroneous constituency results [...] that one cannot rely on any figures from the ECK."***[2]

Response 1 IREC's claim that their findings from a sample of 19 constituencies implies that the rest of the constituency results are equally flawed. This claim cannot be supported with the methods that IREC used, and demonstrates a basic misunderstanding of simple statistical concepts.

The reasoning underpinning IREC's decision not to pursue statistical analysis demonstrates a misunderstanding of basic statistical concepts like random sampling, as well as an ignorance more sophisticated statistical election forensics. Their tenuous logic on why not to do electoral statistics proceeds in two steps. We examine each step in turn, assessing the validity of each step. First, IREC states that "[a]lmost all parliamentary and presidential election results for the constituencies sampled are erroneous, which means that very few of the officially published figures are actually accurate" (p. 127). Put another way, IREC is saying that because the ECK's results in the 19 "sampled" constituencies contained many errors, most of the other 191 constituencies contain similar errors. A student with even the most basic statistical training will recognise this conclusion as false. The mistake that IREC made lies in the way it chose its sample. In short, because IREC's sample focuses on disputed constituencies rather than a random sample of all constituencies, their findings in that sample cannot generalise to all 210 constituencies. Again, we quote the report in full to capture the magnitude of their error:

> The magnitude of problem that brought about the dispute in results from the 2007 elections is enormous. The entire result and a significant part of the election was in serious dispute. Therefore, the choice of what to analyze needed to be very strategic and robust, but at the same time careful enough in order not to leave out key areas that had serious disputes, or those that could be instrumental in helping the IREC achieve its mandate. In order to identify the constituencies for analysis we were guided by a number of criteria, namely:
>
> a. Looking at specific disputed constituencies listed by various parties/bodies. For instance, submissions sent to IREC by Orange Democratic Movement (ODM), Party of National Unity (PNU) and other bodies, listed a number of problematic constituencies: [sic]
>
> b. Selecting areas using a structured selection criteria based on various electoral anomalies such alleged high-turn out and so on [sic]

[2] IREC report, p. 136.

c. Areas with special features or attributes that are prone to electoral abuse, such as the constituencies that were last to submit results, areas where major complaints were raised, and so on. (Annex 6.A, p. 9)

This excerpt clearly states IREC's criteria in choosing constituencies for more detailed analysis: they examined constituencies where problems were most likely to be found. Their choice to focus on problematic constituencies is understandable. However, in doing so, they preclude the possibility of drawing general statistical conclusions about all 210 constituencies. Their sample of constituencies significantly over-represents constituencies that were considered "problematic" after the election. As a result, we cannot conclude, as IREC does, that "very few of the officially published figures are actually accurate."

Demonstrating that IREC's sample is not representative of the 210 constituencies is quite easy. If their sample is representative of all constituencies, then the average (and standard deviation) of constituency-level characteristics in the sample should be very similar the averages of those variables for all 210 constituencies. In table 1, we show that this is simply not the case. Note several key differences. In IREC's sample, Kibaki received, on average 62% of the vote, whereas the average for all constituencies is only 41%.[3] In IREC's sample, constituencies tended to have much lower population densities, and many more registered voters. Perhaps the most telling difference is the percentage difference between presidential and parliamentary votes. If we accept that the difference between those two vote counts is a decent measure of the "material defectiveness" of the ECK's data, then we can clearly see the falsity of IREC's contention that all of the ECK's data is hugely problematic. In IREC's sample, the average percentage difference between presidential and parliamentary votes cast is 7.1%, while the average over all constituencies is a mere 2.5%.

An analogy helps to clarify the problem at hand. Suppose a farmer has 100 chickens, and he wants to estimate about how much he will earn if he takes them to market and sells them. He knows that different sized chickens fetch different prices, and he is not sure how many large and how many small birds he has. Our farmer decides that he's going to catch 10 chickens, and use those 10 to make a generalization about the larger flock. He does so, and finds that the ten birds he caught are rather scrawny, with little meat on their bones. Crest-fallen, he sends the chickens to market with his wife, telling her not to expect too much given the sorry state of their flock. That evening, she returns, her purse bulging with money, and tells her husband that they did quite well at the market with all the large chickens. What had the farmer done wrong when estimating the value of his flock? His "sample" contained the weaker, sicker chickens, since they were easier to catch than healthy chickens. As a result, his estimate of the nature of the flock was not accurate, and he should not have concluded that his flock was full of small birds. Analogously, IREC cannot – based on its sample – conclude that all of the electoral data is of poor quality, because they chose constituencies that were more likely to have problems. Thus, contrary to their claims, we

[3]This does not imply that Kibaki got less than 50% of the vote. Because we are interested in constituency-level characteristics, we weighed all constituencies equally in our averaging: constituencies with a small number of registered voters and constituencies with many registered voters were treated the same, rather than weighted by their size.

Variable	IREC Sample	All Constituencies
Registered Voters, Nov. 2007	80067 (35831)	66884 (33068)
Population Density	449.78 (868.44)	787.30 (2870.88)
% Urban Population	82.3% (25.7%)	86.7% (26.1%)
Poverty Incidence	47.53 (15.27)	53.51 (13.12)
Inequality	34.27 (3.59)	33.25 (3.07)
Change in Registered Voters, May-Nov. '07	8% (7%)	10% (7%)
Kibaki Vote Share, 2007	62.4% (34.6%)	41.3% (33.5%)
% Difference, Pres. and Parl. Votes	7.7% (7.1%)	2.5% (3.1%)
# Polling Stations	91.05 (35.06)	99.08 (42.78)
# Parliamentary Candidates	15.74 (6.27)	11.77 (6.17)
# Civic Candidates	95.74 (52.82)	70.61 (39.17)

Table 1: Comparison of IREC's sample and all constituencies: By comparing the average and standard deviations of 11 variables, we find that there are significant differences in the sample of 19 constituencies that IREC examined and the 210 constituencies as a whole. As a result, IREC cannot convincingly assert that findings in their sample generalize to the rest of the constituencies. Number in parentheses is the relevant standard deviation.

do not know whether or not other constituencies have errors as substantial as those seen in IREC's sample. If IREC had followed proper statistical procedures and chosen a random sample of constituencies, then this knowledge would be within our grasp. Unfortunately, they did not.

Cursory dismissal of statistical modeling

Problem 2 ***"IREC decided that it was not worthwhile to conduct any more profound statistical analysis of the available turnout data etc. The results would be almost useless for IREC's purposes as they could not form a basis for useful conclusions."***[4]

Response 2 *As a result of IREC's erroneous assumption that the results of their sample generalize to the rest of the constituencies, they concluded that more advanced statistical analysis would not be helpful. Even if their assertion were true, statistical tools exist to deal with messy, problematic data like that produced by the ECK. Thus, IREC missed an important opportunity to investigate the data for fraud in a more detailed manner.*

IREC claims that statistical analysis of the ECK data should not be undertaken, because "the official ECK election results (published on the website and elsewhere) have not been cleaned of mistakes of a purely arithmetical nature" (p. 136). As discussed immediately above, we do not know, given IREC's statistical analysis, the state of the electoral data in the other constituencies, since IREC's sample is not representative. Regardless of this fact, the problem of sloppy, partially-incorrect data is not a Kenyan innovation; it is a problem that statisticians have dealt with for years. As a result, a number of statistical methods have been developed to assess this kind of data. The aim of such "robust" methods (as they are called in the academic statistical literature) is two-fold. First, they aim to reduce the influence of anomalous data-points when estimating a statistical relationship between different variables. Second, by uncovering the true relationship between variables, robust methods help separate anomalous data points from "normal" data points. By achieving these two goals, robust methods let us estimate the actual relationship between two variables (e.g., parliamentary vote counts and presidential vote counts), while filtering out the impact of data points containing gross errors.

Applying these methods to electoral data is not a new idea. In fact, statisticians and political scientists have applied these robust methods to electoral data in the past. Most recently, Mebane and Sekhon (2004) developed a method which identified anomalous votes for a third-party candidate in the 2000 US presidential election. In that paper, the authors are able to successfully identify anomalous voting results in the state of Florida – results which led to George Bush's election as president of the United States. Other examples of the application of methods like this to electoral data exist (Wand et al., 2001; Myagkov, Ordeshook and Shakin, 2005) Needless to say, IREC did not attempt to apply this family of methods to Kenyan electoral data. Moreover, the Electoral Commission still has not released the polling centre level results most relevant to such at analysis. Thus, efforts to apply such methods to electoral data remain stymied. As we have shown above, none of IREC's evidence demonstrates that the ECK's electoral data is completely bad; and even if much of it were, statistical models exist to help shed light on where anomalies occurred and who they benefited.

[4]IREC report, p. 136.

Failure to examine statutory forms using appropriate statistical tests

Problem 3 ***"In the case of 2007, where there were vast allegations of changes in statutory forms, this again could very well pass statistical tests, but can only be faulted on account of breaching certain laws or electoral codes and practices."***[5]

Response 3 In fact, statistical tests exist that can detect electoral fraud resulting from changes made in statutory forms. Unlike other approaches that rely on assumptions about past or concurrent voting behavior, or set arbitrary thresholds for "unlikely" voting behavior (like high turnout), these statistical tests rely on the tried-and-true patterns that appear in numbers like vote counts. If an official commits electoral fraud by changing a candidate's vote count on a statutory form, then these tests are likely to detect deviations from that pattern. IREC appears not to have considered these methods nor applied them to data from the 1702 polling centres which they examined in detail.

This family of statistical tests are based on a statistical feature called "Benford's Law," after Frank Benford, an American physicist who noticed that the *digits* of many different kinds of numerical data follow predictable patterns (Benford, 1938). An example helps to illustrate. Imagine that we have numerical data from a set of polling stations in a disputed constituency. That data might look something like this:

Polling Station	Candidate #1	Candidate #2	Candidate #3	Total Votes Cast
Polling Station #1	0	8**2**6	12	8**38**
Polling Station #2	0	1**1**1	38	1**49**
Polling Station #3	31	3**6**3	53	4**47**
Polling Station #4	7	9**8**5	1123	21**15**
Polling Station #5	0	1**8**2	45	2**27**
Polling Station #6	0	2**1**7	0	2**17**
⋮	⋮	⋮	⋮	⋮

Table 2: Statistical Tests for Electoral Fraud: The pattern of digits in lists of numbers (like these hypothetical vote returns) generally follow well-established patterns. By looking for deviations from those patterns, we can detect irregularities that may indicate fraud.

The data contained in table 2 is not unlike that contained in a statutory form 17A. It simply reports the vote counts for each candidate by polling station. Various tests based on Benford's Law exist to test for fraud in data like this. For instance, Mebane (2006) presents a "second-digit" test that looks at how often each of the digits (that is, the numbers zero through nine) appear in a series of vote returns. These digits are underlined in table 2, as

[5] Annex 6.A, p. 9.

an example, for candidate #2. Since Benford's Law provides us with a credible, objective benchmark with which to compare these second-digits, we can look for deviations from that benchmark. Mebane (2006, p. 3) explains: "Benford's Law specifies that in a collection of numbers the different digits should not occur with equal frequency. That is, each of the nine possible first significant digits $(1, 2, ..., 9)$ should not each occur one-ninth of the time, each of the ten possible second significant digits $(0, 1, ..., 9)$ should not each occur one-tenth of the time, and so forth." Table 3 shows the frequency at which these second digits should appear according to Benford's Law.

Digit	0	1	2	3	4	5	6	7	8	9
First	–	.301	.176	.124	.097	.079	.067	.058	.051	.046
Second	.120	.114	.109	.104	.100	.097	.093	.090	.088	.085

Table 3: First- and Second-Digit Frequencies from Benford's Law. Frequencies for third, fourth, and later digits can also be derived. Source: Table 1, Mebane (2006).

Beber and Scacco (2008) takes this examination of the digits contained in electoral returns in another direction, examining the frequencies of the final two digits, as emphasized in the "Total Votes Cast" column of table 2. That research integrates research from psychology which shows that, when generating "random" numbers, human beings produce specific patterns that can be recognized. For instance, humans tend to generate adjacent numbers (e.g., "12" or "87") than chance alone would generate, and avoid repeating numbers (e.g., "44") (Nickerson, 2002). Beber and Scacco (2008) apply these techniques to electoral data from Sweden and Nigeria, finding very little evidence for fraud in the former and significant evidence in the latter.

In the process of their investigations, IREC examined in detail the form 17A's, which contain data like that in table 2. That data was not subjected to these kinds of analysis, which could have helped differentiate results arising from a normal electoral process, results arising simple, unintentional "human error" – like that emerging from mis-transcription or incorrect arithmetic – and results arising from intentional falsification. Unfortunately, even though they examined data from 19 constituencies in detail, they did not apply these simple tests to the data.

Given IREC's reluctance to rely on any source of data as an "objective" benchmark against which to compare numbers reported by ECK or other parties, one would have hoped that they would employ "industry-standard" electoral forensics relying on well-established statistical facts like Benford's Law to examine the veracity of vote counts and turnout numbers at the polling station and form 17A levels. This kind of statistical evidence, combined with a qualitative examination of the inconsistent nature in which many statutory forms (specifically forms 16A and 17A) were filled out, would have provided a much clearer differentiation between fraud and incompetence.

Improper research design relative to the mandate

Problem 4 ***"In the final argument, ODM persisted in contending that, inasmuch as there has been no adequate refutation of such a plot, given the alterations and inconsistencies in the results and the documentation, a finding of fraud is indicated. PNU and the ECK submit that there has been no evidence that any of the alterations and inconsistencies were intended fraudulently to benefit any candidate or in fact had such result. Therefore they contend for a finding that the explanation must be human error. It was not possible to attain consensus as between the members of IREC on this issue. Nor was it necessary. There is indeed consensus in respect to item (e) of the Terms of Reference, relative to the integrity of results, especially in relation to the presidential election: The conduct of the 2007 elections in Kenya was so materially defective that it has been, and will remain, impossible for IREC to establish true and reliable results for the presidential and parliamentary elections."***[6]

Response 4 *In order to make effective recommendations regarding reform of Kenya's electoral system and processes, IREC needed to establish where and why vote counting went wrong. The finding that the results were "materially defective" adds nothing to our knowledge of what went wrong with the ECK – and provides no leverage in terms of what reforms make the most sense. Without trying to find the truth about what went wrong and why, IREC cannot diagnose the specific problems with the ECK.*

At least two problems plague attempts to detect electoral fraud. First, differentiating between "human error" and intentional fraud can be a difficult task. Leveraging several types of evidence on the same area or polling station, however, can go a long way towards telling one from the other. For instance, the application of a Benford's Law-type test to polling station data, in addition to an examination of which candidate benefits from "mistakes" and testimony from party agents, observers, and election officials, should be enough to tell fumble from fraud.

Second, fraud can occur at many different levels, either independently or simultaneously. For instance, a presiding officer at the polling station might falsify electoral returns submitted on a form 16A. A returning officer might do something similar on the constituency-level form 17A. And a supervisor at KICC might shuffle adjust votes between constituencies at the province-level. A suitable research design should be able tell the difference between an honest mistake and intentional fraud, as well as differentiate fraud on one level from that on another.

These requirements for a suitable research design have a practical purpose. Even if we accept IREC's assertion that figuring out who won is not in their mandate, without a suitable research design, IREC would not be able to fulfill another key part of its mandate: to make substantive recommendations on the reform of the Electoral Commission of Kenya. Because it did not develop a cogent approach to understanding what problems happened where and

[6]IREC Report, pp. 125-6.

at what level during the elections, IREC could not effectively differentiate human error from attempts at fraud, nor locate either of these phenomenon at the polling-station, constituency, or national levels. As a result, Kenyans received an report telling them much of what the already know: that the elections contained many problems, including bribery, vote-buying, intimidation, and the like.

IREC's errors in research design may lie at the root of its unnecessarily vague findings. Was IREC's statistical research design capable of adjudicating whether or not there was rigging at KICC or at any other level, for that matter? To do so, their design would have to achieve two goals. First, it would have to differentiate between human error and fraud. In the report, human error is generally associated with a stressful and complex voting environment. IREC states that their "analysis of counting and tallying in ten constituencies with huge turnout discrepancies demonstrates convincingly that the discrepancies are probably due to human error and general incompetence, difficult working conditions at constituency tallying centres, pressure from KICC-based ECK staff on returning officers to provide fast results, pressure from candidates and incumbents eager to know their own electoral fate, lack of training and unclear messages as to when and how erroneous constituency results might be corrected."[7] However, these claims (e.g., general incompetence, difficult working conditions, etc.) are simply theories. If difficult voting environments caused more discrepancies, then discrepancies should be correlated with factors we think cause "difficult voting environments." IREC did not examine these theories using even the most basic statistical tools at the polling station or constituency level.[8]

For example, suppose we believe that having a high number of civic candidates leads to a stressful voting environment, and that a stressful voting environment leads to greater absolute discrepancies between the presidential and parliamentary vote counts. Then, we would expect a positive relationship between the number of civic candidates and the presidential/parliamentary vote discrepancies. A simple correlation between these two variables would be a good first step towards understanding whether or not such a relationship actually exists. IREC did not carry out even these most basic tests on either their sample or the ECK's data in order to validate the theories upon which they hang ECK's shortcomings. While testing for fraud directly may be difficult, testing for human error was within the grasp of IREC. Doing so would have provided some much-lacking evidence for their claims.

The statistical tests based on Benford's Law (described above) provide another way to differentiate between human error and fraud. IREC did not apply these tools to their sample. This is especially problematic, since that class of tests would have provided valid inferences in their sample. Since these tests are based on the frequencies of digits in a list of numbers, and not the relationship between different variables, they provide an external benchmark against which to compare the ECK's data. Given that IREC was quite reluctant to believe any statistical analysis relying on ECK data, it only seems natural that these digit-based

[7] IREC Report, p. 131.

[8] Even if they had done so using data from their sample, those findings would not provide a clear picture of the actual relationship between variables due to the fact that their sample of polling stations was not representative.

tests – commonly used in electoral forensics – be employed in the Kenyan case.[9]

The second flaw in IREC's research design lies in its inability to attribute errors – fraudulent or otherwise – to a specific point in the counting process. Given IREC's reluctance to believe analyses based on ECK data, it seems odd that their "analysis relied only on official documents and results submitted to IREC by ECK."[10] One could argue that, since the ECK may have felt threatened by IREC's mandate, documents coming from ECK could have been manipulated to aid in their absolution with respect to fraud. We have no evidence of this hypothesis. However, if IREC finds other analyses using ECK data unconvincing, why should IREC's own analysis of documents that had been in the possession of the ECK since the elections be credible?

All speculation aside, if we assume that the documents provided to IREC by the ECK are genuine, could their analysis determine whether or not fraud took place at KICC? Again, the answer appears to be "no". A basic point of departure for many criminal investigations is "cui bono?" – who benefits? Unfortunately, IREC's unorthodox sampling procedures prevent any meaningful inference about who may have benefitted from the changes made at KICC, i.e., differences between the results on form 16 and the official ECK final results. Returning to table 1, we see that constituencies in the IREC sample have, on average, a much higher level of support for Kibaki than the average of all 210 constituencies. That is, IREC's sample over-represents Kibaki-supporting areas. IREC seems to understand that their use of this sample to make inferences about who benefitted from these discrepancies is improper: "Even though the sample used was designed with other purposes in mind, it might, however, be of interest to conduct a brief analysis of the impact of the errors on the numbers of votes obtained by the main candidates.[...] What the table [of discrepancies by constituency] clearly shows is the random nature of the errors in these constituencies: no candidate can be said to have benefited."[11] A more responsible presentation of these results would have stated the over-representation of Kibaki-supporting constituencies in the sample, a fact that could plausibly explain why it appears that no candidate systematically benefitted from discrepancies between the IREC and ECK numbers.

Moreover, IREC did not recognize ECK's opportunity to commit a kind of fraud at KICC that is uniquely different than fraud occurring at lower levels. Only at the national tallying centre could a coordinated transfer of votes between constituencies, into rejected votes, or between candidates been carried out. In order to detect such subtle changes, IREC would have had to examine the results of an entire province or even multiple provinces, a task they were clearly unwilling to undertake.

[9] See IREC Report, p. 137, on IREC's reluctance to rely on analyses using ECK data.
[10] IREC Statistical Annex 6.A, p. 9
[11] IREC Report, pp. 129-130.

Conclusion

A few simple changes in IREC's research design would have enabled them to diagnose the various problems that occurred, without a significant increase in the cost or effort required. For instance, a focus on form 17A's and 16's in comparison with ECK's final figures (for either all 210 constituencies or a substantial random sample) would have addressed the issue of rigging at KICC and as well as provided information about the veracity of constituency-level results. This approach would have allowed us to detect indications of fraud at the polling station and constituency-level, though not differentiate between the two, since vote re-counts would be required to verify the results of a given polling-station. In addition, the approach would have allowed a clearer understanding of exactly how changes made at KICC affected the outcome, after correcting for human arithmetic error on the form 17A's. This approach would have been superior to the one chosen by IREC, in terms of both diagnosing problems within the structure of ECK (presiding/returning officer, KICC, etc.) and the spatial location of likely fraud. And, this approach would have obviated the need for time- and effort-consuming re-tallies from the form 16A's at the constituency level. These criticisms not withstanding, engaging a document management firm to re-tally all form 16A's was likely well within the budget and timeframe of IREC, and would have provided the most comprehensive understanding of how and where fraud and human error affected the results of the 2007 General Elections.

IREC's statistical treatment of data from the 2007 General Election suffers from many flaws. Basic choices in their approach – specifically their non-random sampling procedure – make inferences about general patterns in the constituency-level results untenable, even though IREC made such unsupported claims repeatedly in their report. By not using more advanced statistical models and tests on the ECK data at the polling station and constituency levels, IREC essentially left relevant and useful evidence on the table, unexplored. Better preparation, a well-designed approach to the question of fraud, and the use of the correct tools for electoral forensics would have resulted in a more credible – and more useful – report.

References

Beber, Bernd and Alexandra Scacco. 2008. "Identifying Electoral Fraud: A Novel Test and New Data from Nigeria." *Columbia University, Department of Political Science Working Paper* .

Benford, Frank. 1938. "The Law of Anomalous Numbers." *Proceedings of the American Philosophical Society* 78(4):551–572.

Mebane, Walter R. 2006. "Election Forensics: Vote Counts and Benford's Law." *Working Paper* .

Mebane, Walter R. and Jasjeet S. Sekhon. 2004. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science* 48(2):392–411.

Myagkov, Mikhail, Peter Ordeshook and Dimitry Shakin. 2005. "Fraud or Fairytales: Russia and Ukraine's Electoral Experience." *Post-Soviet Affairs* 21(2):91–131.

Nickerson, Raymond S. 2002. "Perception and Production of Randomness." *Psychological Review* 109(2):330–357.

Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Michael C. Herron and Henry E. Brady. 2001. "The Butteryfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95(4):793–810.