Auditing Australian Senate Ballots

Berj Chilingirian[*1], Zara Perumal[1], Ronald L. Rivest[1],
Grahame Bowland[†2], Andrew Conway[‡3], Philip B. Stark[4],
Michelle Blom[5], Chris Culnane[5], and Vanessa Teague[5]

[1]Computer Science and Artificial Intelligence Laboratory,
Massachusetts Institute of Technology.
`[berjc,zperumal,rivest]@mit.edu`
[2]`erinaceous.io` , `grahame@angrygoats.net`
[3]Silicon Econometrics Pty. Ltd., `andrewsa@greatcactus.org`
[4]Department of Statistics, University of California, Berkeley.
`stark@stat.berkeley.edu`
[5]Department of Computing and Information Systems, University
of Melbourne.
`[michelle.blom,christopher.culnane,vjteague]@unimelb.edu.au`

November 8, 2016

Abstract

We explain why the AEC should perform an audit of the paper Senate ballots against the published preference data files. We suggest four different post-election audit methods appropriate for Australian Senate elections. We have developed prototype code for all of them and tested it on preference data from the 2016 election.

---

[*]Authors are grouped by institution, in alphabetical order, and then listed in alphabetical order within each institution.

[†]Grahame Bowland is a member of the Australian Greens. His contribution to this project has consisted entirely of help in implementing the Australian Senate counting rules and facilitating Bayesian audits using his code. The techniques here are non-political.

[‡]Andrew Conway is a member of the Secular Party.

1

Contents

1 Introduction

A vote in the Australian Senate is a list of handwritten numbers indicating preferences for candidates. Voters typically list about six preferences, but may list any number from one to more than 200. Ballots are scanned, digitized and then counted electronically using the Single Transferable Vote (STV) algorithm [Aus16].

Automating the scanning and counting of Senate votes is a good idea. However, we need to update our notion of "scrutiny" when so much of the process is electronic. We suggest that, when the preference data file for a state is published, there should be a statistical audit of a random sample of paper ballots. This should be performed in an open and transparent manner, in front of scrutineers.

Election outcomes must be accompanied by evidence that they accurately reflect the will of the voters. At the very least, the system should be *Software Independent* [Riv08].

> A voting system is *software independent* if an undetected change or error in its software cannot cause an undetectable change or error in an election outcome.

This principle was articulated after security analyses of electronic voting machines in the USA showed that the systems were insecure [FHF06, KSRW04, BEH+08, CAt07]. The researchers found opportunities for widespread vote manipulation that could remain hidden, even from well-intentioned electoral officials who did their best to secure the systems.

Followup research in Australia has shown election software, like any other software, to be prone to errors and security problems [HT15, CBNT]. For this reason, evidence of an accurate Senate outcome needs to be derived directly from the paper ballots.

Legislation around the scrutiny of the count has not kept pace with the technology and processes deployed to perform the count. As a result, the scrutineering has lost a significant portion of its value. With the adoption of a new counting process the scrutineering procedures need to be updated to target different aspects of the system. The current approach might comply with legislation, but it doesn't give scrutineers evidence that the output is correct.

This paper suggests four different techniques for auditing the paper Senate ballots to check the accuracy of the published preference data files. The techniques vary in their assumptions, the amount of work involved, and the confidence that can be obtained.

These suggestions might be useful in two contexts:

- if there is a challenge to this year's Senate outcome,

- as an AEC investigation of options for future elections.

An audit should generate evidence that the election result is accurate, or detect that there has been a problem, in time for it to be corrected. We hope that these audits become a standard part of Australian election conduct.

1.1 Q & A

- **Q: Why do post-election audits?**
  A: to derive confidence in the accuracy of the preference data files, or to find errors in time to correct them.

- **Q: What can a post-election audit tell you about the election?**
  A: It can tell you with some confidence that the outcome is correct, or it can tell you that the error rate is high enough to warrant a careful re-examination of all the ballots.

- **Q: Can the conclusion of the audit be wrong?**
  A: Yes, with small probability an audit can confirm an outcome that is, in fact, wrong. It can also raise an alarm about a large error rate, even if the errors do not in fact make the outcome wrong.

- **Q: Who does post-election audits now?**
  A: Many US states require by law, and routinely conduct, post-election audits of voter-verified paper votes when the tallies are conducted electronically. Exact regulations vary—the best examples are the Risk-Limiting Audits [BFG+12] conducted by California and Colorado.

- **Q: What is needed to do a post-election audit?**
  A: The audit begins with the electronic list of ballots, and (usually) relies on being able to retrieve the paper ballot corresponding to a particular randomly-chosen vote in the file. There must also be time and people to retrieve the paper ballots and reconcile them with the preference data file. A video of random ballot selection is here: `https://www.youtube.com/watch?v=sdWL8Unz5kM`.

- **Q: How long does it take? How many ballots must be examined?**
  A: It depends on the audit method, the level of confidence derived, the size of the electoral margin and the number of errors in the sample. This is described carefully below.

- **What is the difference between a statistical post-election audit and a recount?**
  A: It's not feasible to do manual recounts; a statistical post-election audit would provide a comparable way of assessing the accuracy of the outcome.

1.2 Our contribution

This paper describes four suggested approaches to auditing the paper evidence of Australian Senate votes, each described in more detail in Section 2.

Section 2.1 Bayesian audits [RS12],

Section 2.2 a "negative" audit based on an upper bound on the margin,

Section 2.3 a simple scheme with a fixed sample size,

Section 2.4 a "conditional" risk-limiting audit, which tests one particular alternative election outcome.

We have prototype code available for completing any of the above kinds of audit. This would be the first time these sort of auditing steps are being applied, and so this year's efforts would be much more "exploratory" in character than "authoritative". We hope to be able to perform two or more kinds of audits on the same samples. However, we do not even know, at the time of writing, whether any audit will happen at all.

The key objective is to provide evidence that the announced election outcome is right, or, if it is wrong, to find out early enough to correct it by careful inspection of the paper evidence.

1.3 Where the Senate count depends on trusting software

This very brief security analysis of the current process is based on documents on the AEC's website [AEC16]. The objective of the system is, in principle, extremely simple: capture the vote preferences from the ballot papers, and then publish and tally them.

The current implementation results in a number of points of trust, in which the integrity of the data is not checked by humans and is dependent on the secure and error-free operation of the software. Whilst internal audit steps are useful, there are many systematic errors and security problems they would not detect. We list the three most obvious examples below.

Image Scanning There appears to be no verification that the scanned image is an accurate representation of the paper ballot. As such, a malicious, or buggy, component could alter or reuse a scanned image, which would then be utilised for both the automatic and manual data entry. This would pass all subsequent scrutiny, whilst not being an accurate representation of the paper ballot. We understand that scrutineers can ask to see the paper ballot, but this seems very unlikely to happen if the image is clear and the preferences match.

Ballot Data Storage Whilst a cryptographic signature is produced at the end of the scanning and processing stage, and prior to submission to the counting system, this signature is based on whatever is in the database. There is no verification that the database accurately represents what was produced by the automatic recognition or the manual operator, nor that it was the same thing displayed to scrutineers on the screen. An error, or malicious component, with access to the database could undetectably alter the contents.

Signature Checking Automatic signature generation is a problem in the presence of a misbehaving device. There is no restriction on the device creating signatures on alternative data. Likewise, there appears to be no scrutiny over the data being sent between the scanning process and the counting process, particularly, that the sets of data are equal. There appears to be logging emanating from both services, but no clear description of how such logs will be reconciled and independently scrutinised.

In summary, there are plenty of opportunities for accidental or deliberate software problems to cause a discrepancy between the preference files and the paper votes. This is why the paper ballots should be audited when the preference files are published.

1.4 Background on audits

The audit process begins with the electronic data file that describes full preferences for all votes in a state. This file implies a *reported election outcome R*, which is a set of winning candidates which we assume to be properly computed from the preferences in the data file. (Actually we don't have to assume—we can check by rerunning the electronic count.) Each line in the data file is a *reported vote*—we denote them $r_1, \ldots, r_n$, where n is the total number of voters in the state. Each reported vote r_i (including blank or informal ones) corresponds to an *actual vote a_i* expressed on paper, which can be retrieved to check whether it matches r_i. The whole collection of actual votes implies an *actual election outcome A*. We want to know whether $A = R$.

The audit proceeds by retrieving and inspecting a random sample of paper ballots. A *comparison* audit chooses random votes from the electronic data file and compares each one with its corresponding paper ballot. The auditor records discrepancies between the paper and electronic votes. A *ballot polling* audit chooses paper ballots at random and records the votes, without using the electronic vote data.

Although the security of paper ballot processing is important, it's independent of the audit we describe here. An audit checks whether the electronic result accurately reflects the paper evidence. Of course if the paper evidence wasn't properly secured, that won't be detected by this process. Our definition of "correct" is "matching the retained paper votes."

An election audit is an attempt to test the hypothesis "That the reported election outcome is incorrect," that is, that $R \neq A$. There are two kinds of wrong answer: an audit may declare that the official election outcome is correct when in fact it is wrong, or it may declare that the official outcome is wrong when in fact it is correct. The latter problem is easily solved in simpler contexts by never declaring an election outcome wrong, but instead declaring that a full manual recount is required. The first problem, of mistakenly declaring an election outcome correct when it is not, is the main concern of this paper.

An audit is *Risk-limiting* [LS12] if it guarantees an upper bound on the probability of mistakenly declaring a wrong outcome correct. A full manual recount is risk-limiting, but prohibitively expensive in our setting. None of the audits suggested in this paper is proven to be risk limiting, however all of them provide some way of estimating the rate of errors and hence the likelihood that the announced outcome is wrong. In some cases, the audit may not say conclusively whether the error rate is large enough to call the election result into question. In others, we can derive some confidence either that the announced outcome is correct or that a manual inspection of all ballots is warranted.

1.5 Why auditing the Australian Senate is hard

Election auditing is well understood for US-style first-past-the-post elections but difficult for complex voting schemes. The Australian Senate uses the Single Transferable Vote (STV). There are many characteristics that make auditing challenging:

- **It is hard to compute how many votes it takes to change the outcome.** Calculating winning margins for STV is NP-hard in general [Xia12], and the parameters of Australian elections (sometimes more

than 150 candidates) make exact solutions infeasible in practice. There are not even efficient methods for reliably computing good bounds.

- **A full hand count is infeasible,** since there are sometimes millions of votes in one constituency,

- **In practice the margins can sometimes be remarkably small.** For example, in Western Australia in 2013 a single lost box of ballots was found to be enough to change the election outcome. In Tasmania in 2016 there were more than 300,000 votes, but the final seat was determined by a difference of 141 votes (meaning errors in the interpretation of 71 ballots might have altered the outcome).

This makes it difficult to use existing post-election auditing methods.

To get an idea of the fiendish complexity of Australian Senate outcomes, consider the case of the last seat allocated to the State of Victoria in 2013. Ricky Muir from the Australian Motoring Enthusiasts Party won the seat, in a surprise result that ousted sitting Senator Helen Kroger of the Liberal party. In the last elimination round (round 291), Muir had 51,758 more votes than Kroger, and this was generally reported in the media as the amount by which he won. However, the true margin was less than 3000 (about 0.1%). If Kroger had persuaded 1294 of her voters, and 1301 of Janet Rice (Greens)'s voters, to vote instead for Joe Zammit (Australian Fishing and Lifestyle Party), this would have prevented Zammit from being excluded in count 224. Muir, deprived of Zammit's preferences, would have been excluded in the next count, and Kroger would have won. (Our algorithm for searching for these small margins is described in the full version of this paper.)

This change could be made by altering 2595 ballots, in each case swapping two preferences, none of them first preferences, all below the line. First preferences are relatively well scrutinised in pollsite processes before dispatch to the central counting station. Other preferences are not. Also *lowering* a particular candidate's preference wouldn't usually be expected to help that candidate (though we are not the first to notice STV's nonmonotonicity). So the outcome could have been changed by swapping poorly-scrutinised preferences, half of which seemed to disadvantage the candidate they actually helped, in far fewer ballots than generally expected.

2   Overview of available options

This section describes four different proposals and compares them according to the degree of confidence derived, the amount of auditing required, and other assumptions they need to make. We have already implemented prototype software for running Bayesian Audits (Section 2.1) and computing upper bounds on the winning margin (Section 2.2). We have tested the code on the AEC's full preference data from some states in the 2016 election—results are described briefly below.

2.1   Bayesian Audits

Rivest and Shen's "Bayesian audit" [RS12] evaluates the accuracy of an announced election outcome without needing to know the electoral margin. It

samples from the posterior distribution over profiles of cast ballots, given a prior and given a sample of the cast paper ballots (interpreted by hand). It only looks at a sample of the cast paper ballots—it does not compare the sampled paper ballots with an electronic interpretation of them.

An *profile* is a set of ballots. The auditor doesn't know the profile of cast (paper) ballots, and so he works with a probability distribution p over possible such profiles, which summarises everything the auditor belives about what the profile of cast ballots may be.

The Bayesian audit proceeds in stages. Successive stages consider increasingly larger samples of the cast ballots.

Each stage of the Bayesian audit provides an answer to the question "what is the probability of various election outcomes (including the announced outcome), if we were to examine the complete profile of all cast ballots?"

This question is answered by simulating elections on profiles chosen according to the posterior distribution based on p, and measuring the frequency of each outcome.

Each audit stage has three phases:

1. audit some randomly chosen paper ballots (that is, obtain their interpretations by a human),

2. update p using Bayes' Rule,

3. sample from the posterior distribution on profiles determined by p and determine the election outcome for each; measure the frequency of different outcomes.

Like any process that uses Bayes' Rule, choosing a prior is a key part of the initialization. The suggestion in [RS12] is to allow any political partisan to choose the prior that most supports their political beliefs. When everyone (who uses Bayes' Rule properly) is satisfied that the evidence points to the accuracy of the announced result, the audit can stop. For example, the auditors could agree to stop when 95% of simulated election outcomes match the reported outcome.

In the Australian Senate case, we assume that there will be only one apolitical auditing team (though in future candidate-appointed scrutineers could do the calculations themselves). Hence we suggest a prior that is neutral—if the announced outcome is correct, this probability distribution will be gradually corrected towards it.

An alternative, simpler version amounts to a bootstrap, treating the population of reported ballots as if it is the (prior) probability distribution of ballots, and then seeing how often one gets the same result for samples drawn from that prior. This gives an approximate indication of how much auditing of paper ballots would be necessary, assuming that the paper ballots were very similar to the electronic votes. We have run this version of the audit on the Senate outcome from 2016. Table 1 shows the number of samples needed in the bootstrapping version, in order to get 95% of trials to match the official outcome. Tasmania is the closest, and the only one that's really infeasible: a sample size of about 250,000 ballots is needed before 95% of trials produce the official outcome, which is not much better than a complete re-examination of all ballots. This is hardly surprising given the closeness of the result. Queensland requires 23,000, which is still only a tiny fraction of the total ballots. Apart from that, all the other states require only a few thousand samples.

State	Number of votes (millions)	Audit sample size (thousands)
NSW	4.4	4.6
NT	0.1	1.5
Qld	2.7	23
SA	1.1	3
Tas	0.34	250
Vic	3.5	6
WA	1.4	9

Table 1: Sample sizes for 95% agreement in bootstrap Bayesian Audit.

We suggest a combination of the bootstrapping method with the retrieval of paper ballots: have a single short partial ballot in favor of each candidate, combined with an empirical Bayes approach that specifies that only ballots of the forms already seen in the sample (or the short singleton ballots) may appear in the posterior distribution.

Although these audits were designed for complex elections, there are significant challenges to adapting them to the Australian Senate. Running the simulations efficiently is challenging when the count itself takes some time to run. Answers to these challenges are described in the full version of the paper.

2.2 Upper bounds on the margin plus "negative" audits

We have implemented some efficient heuristics for searching for ways to change the election outcome by altering only a small number of votes—the code is available at `https://github.com/SiliconEconometrics/PublicService`. The Kroger/Muir margin described in the Introduction is an example. We can guarantee that the solution we find is genuine, *i.e.* a true way to change the outcome with that number of ballots, but we can't guarantee that it is minimal—there might be an even smaller margin that remains unknown. The algorithm produces a list of alternative outcomes together with an upper bound on the number of votes that need to change to produce them.

If the error rate is demonstrably higher than this upper bound on the margin, then we can be confident it is large enough to change the election result. Of course, it does not follow that the election result is wrong, especially if the errors are random rather than systematic or malicious. It means that all the paper evidence must be inspected.

This allows a "negative audit," which can allow us to infer with high confidence that the number of errors is high enough.

Suppose there are N ballots in all. Suppose we know that the outcome could be altered by altering no more than X ballots in all, provided those ballots were suitably chosen. Suppose we think the true ballot error rate p (ballots with errors divided by total ballots, no matter how many errors each ballot has) is q, with $qN \gg X$; that is, we think the error rate is large enough that the outcome could easily be wrong. Then a modest sample of size n should let us infer with high confidence that $pN > X$.

For example, consider the 2016 Tasmanian Senate result, in which the final margin was 71 out of 339,159 votes (a difference of 141 votes). We can compute the confidence bounds based on a binomial distribution. A lower 95% confidence

bound for p if we find 3 ballots with errors in a sample of size 2500 is about 0.0003. That's much greater than the error rate of $71/339,159 = 0.00021$ that would be needed to change the outcome. If we did find errors at about that rate, it would be strong evidence that a full re-examination of all the paper ballots is warranted. Code for this and other probability computations in this paper is available at `https://gist.github.com/pbstark/58653bbc26f269d4588ea7cd5b2e12bf`.

2.3 Audits of fixed sample size

A much simpler alternative is to take a fixed sample size of paper ballots (e.g. 0.1% of the cast ballots), draw that many ballots at random and examine them all.

This conveniently puts a "cap" on the number of randomly-chosen paper ballots to be examined, but the audit results may provide less certainty than an uncapped audit would provide.

2.3.1 Risk-measuring audits

Assume now that the aim is to try to find confidence that the election outcome is correct. This audit could quantify the confidence in that assertion, by computing binomial upper confidence bounds on the overall error rate. The idea is to find the p-value (or confidence level) that the sample you actually have gives you that the outcome is right.

Even an error rate of 0.0002, *i.e.*, two ballots with errors per 10,000 ballots, could have changed the electoral result in Tasmania, depending on the exact nature of those errors. The sample size required to show that the error rate is below that threshold—if it is indeed below that threshold—is prohibitively large. If we take a sample of 1,000 ballots and we find no errors that affect the 71 margin, the measured risk is the chance of seeing no errors if the true error rate is 0.0002, *i.e.*, $(0.0002)^0 * (1 - 0.0002)^{1000} = 81\%$. If we took a sample of 2,000, the measured risk would be $(0.0002)^0 * (1 - 0.0002)^{2000} = 67\%$.

However, this method might be quite informative for other contests. Manual inspection of a sample of 1,000 ballots could give 99% confidence that the error rate is below 0.0046 (46 ballots with errors per 10,000 ballots), if the inspection finds no errors at all. If it finds one ballot with an error, there would be 99% confidence that the error rate is below about 0.0066 (66 ballots with errors per 10,000 ballots).

Similarly, manual inspection of a sample of 500 ballots could give 99% confidence that the error rate is below 0.0092 (92 ballots with errors per 10,000 ballots), if the inspection finds no errors at all. If it finds one ballot with an error, there would be 99% confidence that the error rate is below about 0.0132 (132 ballots with errors per 10,000 ballots).

If more errors are found, this gives a way to estimate the error rate. If it is large, this would give a strong argument for larger audits in the future.

2.3.2 Fixed-size samples with Bayesian Auditing

We can also derive some partial confidence measures from the given sample. For example, you could list, for each candidate, the precentage of the time that candidate was elected across the Bayesian experiments. (Each experiment starts

with a small urn filled with the 14000 ballots, plus perhaps some prior ballots, and expands it out to a full-sized profile of 14M ballots with a polya's urn method or equivalent. This is for a nationwide election; for the senate the full-size profiles are the size of each senate district.) Depending on the computation time involved, we might run say 100 such experiments. So, you might have a final output that says:

Joe Jones	99.1 %
Bob Smith	96.2 %
Lila Bean	82.1 %
...	
Rob Meek	2.1 %
Sandy Slip	0.4 %
Sara Tune	0.0 %

Such results are meaningful at a human level, and show what can be reasonably concluded from the small sample.

This allows us to have a commitment to a given level of audit effort, rather than a commitment to a given level of audit assurance, and then give results that say something about the assurance obtained for that level of effort.

2.4 Conditional Risk Limiting Audits

Back to the Tasmanian 2016 example again. One way to examine the issue is to consider the particular, most obvious, alternative hypothesis, *i.e.* that the correct election result differs only in changing the final tallies of the last two candidates. If we assume that all the other, earlier, elimination and seating orders are correct, we can conduct a risk-limiting audit that tests only for the one particular alternative hypothesis. (Of course, it isn't truly risk limiting because it doesn't limit the risks of other hypotheses.) This may be relevant in a legal context in which a challenging candidate asserts a particular alternative. This method would provide evidence that the error rate is small enough to preclude that alternative (if indeed it is), without considering other alternatives.

This can be run as a ballot-level comparison audit, in which the electronic ballot record is directly compared with its paper source. When an error is detected, its impact on the final margin can be quantified (a computationally infeasible problem when considering all possible alternative outcomes). A risk-limiting audit could be based on the Kaplan-Markov method from [Sta08]. It allows the sample to continue to expand if errors are found: that is, it involves sequential testing. At 1% risk limit, the method requires an initial sample size of about (10/margin), where the margin is expressed as a fraction of the total ballots cast. Here, that's about 0.0002. A risk limit of 5% would require hand inspection of roughly 16,000 ballots, assuming no errors were found.

2.5 Summary

These four different audit methods could each be conducted on the same dataset. We would generate the sample by choosing random elements of the official preference data file, then fetching the corresponding paper ballot. The Bayesian Audit and the simple capped scheme would then simply treat the paper ballots as the random sample. The upper-bounds based scheme and the conditional risk limiting audiit would consider the errors relative to what had been reported.

There are important details in exactly how the audit is conducted. We suggest that the auditors not see the electronic vote before they are asked to digitize the paper—otherwise they are likely to be biased to agree. However, we also suggest that they are notified in the case of a discrepancy and asked to double-check their result—this should increase the accuracy of the audit itself. Details of this process are interesting future work. It is, of course, important that the audit itself should be software independent.

If the rate of error is high then a high level of auditing is required. With few or no errors, our best estimates of the necessary sample size for each technique applied to the Tasmanian 2016 Senate are:

- for Bayesian audits, about 250,000 samples until 95% of trials match the official outcome,

- for "negative" audits, a sample that found 3 or more errors out of 2500 ballots would give a 95% confidence bound on the error rate (being big enough),

- a fixed sample size of 500 or 1000, even with no errors, seems unlikely to be large enough to infer anything meaningful for Tasmania 2016, though it may be useful for other contexts,

- a conditional risk-limiting audit would require about 16,000 ballots for a risk limit of 5%, assuming no errors were found.

Most other states would probably be easier to audit as they do not seem to be as close.

3 Implementation Summary

All the tools necessary for conducting a Bayesian audit of Australian Senate votes are available as a Python package at `https://pypi.python.org/pypi/aus-senate-audit`, with code and instructions at `https://github.com/berjc/aus-senate-audit`.

Code for searching for small successful manipulations is at `https://github.com/SiliconEconometrics/F`

Code for computing relevant statistical bounds is at `https://gist.github.com/pbstark/58653bbc26f26`

4 Conclusion

Elections must come with evidence that the results are correct. This work contributes some techniques for producing such evidence for the partly-automated Australian Senate count.

All of the audits discussed here can be conducted immediately, using code already available or specifically produced as a prototype for this project.

4.1 Future Work

In the future we could expand the precision with which we record errors and make inferences about their implications. We are also pursuing an easier user interface for administering the audit.

References

[AEC16]   `http://www.aec.gov.au/elections/candidates/files/counting/css-faqs.pdf`,
           http://www.aec.gov.au/elections/candidates/files/counting/css-
           technical-aspects.pdf, http://www.aec.gov.au/elections/candidates/files/counting/senate-
           count-process-diagram.pdf, 2016.

[Aus16]   Australian   Electoral   Commission.   Count-
           ing   the   votes   for   the   senate,   2016.
           `http://www.aec.gov.au/voting/counting/senate_count.htm`.

[BEH+08]   Kevin RB Butler, William Enck, Harri Hursti, Stephen E McLaugh-
           lin, Patrick Traynor, and Patrick McDaniel. Systemic issues in the
           hart intercivic and premier voting systems: Reflections on project
           everest. *EVT*, 8:1–14, 2008.

[BFG+12]   Jennie   Bretschneider,   Sean   Flaherty,   Susannah   Good-
           man,   Mark   Halvorson,   Roger   Johnston,   Mark   Linde-
           man,   Ronald   L   Rivest,   Pam   Smith,   and   Philip   B
           Stark.   Risk-limiting   post-election   audits: Why   and   how.
           `http://www.stat.berkeley.edu/~stark/Preprints/RLAwhitepaper12.pdf`,
           2012.

[CAt07]   California   top   to   bottom   review   of   voting.
           `http://www.sos.ca.gov/elections/voting-systems/oversight/top-bottom-review`,
           2007.

[CBNT]   Andrew   Conway,   Michelle   Blom,   Lee   Naish,   and   Vanessa
           Teague. An analysis of new south wales electronic vote counting.
           `https://siliconeconometrics.github.io/PublicService/CountVotes/NSWLGE2012Million`

[FHF06]   Ariel J Feldman, J Alex Halderman, and Edward W Felten. Security
           analysis of the diebold accuvote-ts voting machine. 2006.

[HT15]   J Alex Halderman and Vanessa Teague. The new south wales iVote
           system: Security failures and verification flaws in a live online elec-
           tion. In *International Conference on E-Voting and Identity*, pages
           35–53. Springer, 2015.

[KSRW04]   Tadayoshi Kohno, Adam Stubblefield, Aviel D Rubin, and Dan S
           Wallach. Analysis of an electronic voting system. In *Security and
           Privacy, 2004. Proceedings. 2004 IEEE Symposium on*, pages 27–40.
           IEEE, 2004.

[LS12]   M. Lindeman and P.B. Stark. A gentle introduction to risk-limiting
           audits. *IEEE Security and Privacy*, 10:42–49, 2012.

[Riv08]   Ronald L Rivest. On the notion of 'software independence'in
           voting systems. *Philosophical Transactions of the Royal Society
           of London A: Mathematical, Physical and Engineering Sciences*,
           366(1881):3759–3767, 2008.

[RS12]   Ronald L Rivest and Emily Shen. A bayesian method for auditing
           elections. In *EVT/WOTE*, 2012.

[Sta08]   P.B. Stark. Conservative statistical post-election audits. *Annals of Applied Statistics*, 2008.

[Xia12]   L. Xia. Computing the margin of victory for various voting rules. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 982—999, 2012.