

1.

Abhandlungen

Vorsicht, Abstimmungsdetektive. Wahlforensik in der direkten Demokratie

Daniel Bochslers¹

Abstract

In der Wahlforschung finden statistische Verfahren, die auf Auffälligkeiten in Wahlresultaten hindeuten, Anklang. Dieser Beitrag diskutiert die Anwendbarkeit solcher Verfahren aus der Wahlforensik für die direkte Demokratie. Er zeigt einige Anwendungen, wo statistische Tests auf mutmassliche Fälschungen von Abstimmungsergebnissen hindeuten. Und er diskutiert die Grenzen der Wahlforensik: viele der Tests rapportieren natürlich auftretende Auffälligkeiten in Abstimmungsergebnissen, etwa aufgrund der Politgeographie oder aufgrund von strategischem Akteursverhalten, als Unregelmässigkeit. Daher sind sie nicht dafür geeignet menschliches Fehlverhalten von natürlichen Phänomenen zu unterscheiden. Und selbst ausgeklügelte statistische Tests beleuchten immer nur einen spezifischen Schritt des Verfahrens und ein ganz konkretes Problem. Daher kann die Wahlforensik bestenfalls Hinweise auf mögliche Unregelmässigkeiten liefern, wird aber nie feststellen können, ob eine Abstimmung fair verlief.

A. Einleitung

Die mehrfach widerlegten Betrugsvorwürfe bezüglich der Wahl Joe Bidens zum US-Präsidenten 2020 werfen einen langen Schatten auf die US-Demokratie. Infolgedessen wächst auch das Interesse in der politischen Praxis für wissenschaftliche Verfahren aus der Wahlforensik, also statistische Verfah-

1 Der Schreibende hat für die Venezianische Kommission des Europarates einen interdisziplinären Bericht über statistische Verfahren der Wahlforensik mitverfasst, und ist Autor der im Text erwähnten Software. Der Autor bedankt sich bei Nadja Braun Binder für hilfreiche Kommentare.

ren, die auf Auffälligkeiten in Wahlergebnissen hindeuten, die möglicherweise auf Unregelmäßigkeiten im Ablauf der Wahl oder in der Auszählung der Stimmen hinweisen.

Weder Wahlbetrug noch die Wahlforensik sind auf Wahlen beschränkt; der Verdacht von Manipulationen fällt auch auf direktdemokratische Abstimmungen, seien es Sachabstimmungen, Unabhängigkeits- oder Verfassungsabstimmungen oder Referenden zur Amtsenthebung gewählter Politikerinnen und Politiker.

Zweifelhafte Praxen in der direkten Demokratie gibt es zuhauf, europäische Beispiele reichen von Napoleon Bonaparte über die ukrainische Krim bis zum jüngsten Verfassungsreferendum in Serbien. Napoleon legitimierte seine Politik mehrfach in Referenden, allerdings kam die Zustimmung vor allem durch autoritäre Mobilisierung und Kontrolle zustande.² Im Referendum zur Gründung des modernen Schweizer Bundesstaates, 1848, kam die zustimmende Mehrheit im Kanton Luzern nur dadurch zustande, dass der Kanton die Nichtstimmenden als Zustimmung wertete.³ Diverse Referenden in Venezuela in den 2000er-Jahren waren durch Einschüchterung und illegitime Mittel zur Wählermobilisierung gekennzeichnet.⁴ Die Volksbefragung von 2014, die die Annexion der ukrainischen Krim durch Russland legitimieren sollte, wird heute gerne als „international nicht anerkanntes Referendum“ betitelt. Dabei geht unter, dass in Präsenz von irregulären Truppen im Parlamentsgebäude und unter Waffengewalt angesetzt wurde, und dass Bewaffnete auch in den Wahllokalen vor Ort waren.⁵ Im serbischen Verfassungsreferendum über eine Justizreform, 2022, setzte die Regierungspartei alle ihre Hebel in Bewegung, um eine respektable Stimmbeteiligung zu garantieren, dazu gehörte Druck auf Angestellte im öffentlichen Dienst und deren Familien und eine klientelistische Kampagne mit sozial benachteiligten Bürgerinnen und Bürgern als Zielgruppe. Verbreitet sind auch „parallele Wählerlisten“, also Listen auf denen entweder Mitglie-

2 M. Qvortrup, *Referendums and Ethnic Conflict* (Philadelphia (PE): University of Pennsylvania Press), 2014, S. 18.

3 A. Kley, Die Bundesverfassung von 1848, *Historisches Lexikon der Schweiz*, 2011, <https://hls-dhs-dss.ch/de/articles/009811/2011-05-03/> (Zugriff 10.6.2022).

4 Á. E. Álvarez, Venezuela 2007: Los motores del socialismo se alimentan con petróleo, *Revista de Ciencia Política*, 2007, S. 265.

5 A. Wilson, *Ukraine Crisis. What it Means for the West* (New Haven/London: Yale University Press), 2014, S. 110.

der des Wahlausschusses oder Parteiaktivistinnen und -aktivisten kontrollieren ob die klientelistisch mobilisierten Bürgerinnen und Bürger tatsächlich am Referendum teilnahmen.⁶

Anders als die Literatur zur Wahlforensik gibt es nur wenig Literatur, die sich mit Wahlfälschungen in der direkten Demokratie und mit statistischen Methoden zur Identifizierung derselben beschäftigt. Dieser Beitrag bietet einen kleinen Überblick darüber. Im ersten Abschnitt geht er auf die Frage ein, wann und warum es in der Praxis der direkten Demokratie zu Manipulationen kommen dürfte, gefolgt von mehreren Abschnitten zur Forensik. Die Methode der Wahlforensik diskutiere ich anhand eines Anwendungsbeispiels aus Russland, gefolgt von Bemerkungen zur Methodologie. Anschließend analysiere ich zwei Typen von Ansätzen aus der Wahlforensik: Ansätze beruhend auf Verteilungen und Zahlentheorie und andere beruhend auf der Wahlgeographie.

B. Wahlfälschung und direkte Demokratie

Nach dem Ende des kalten Kriegs wurden weltweit Wahlen – meistens mit mehreren Parteien – zur Norm, auch in Ländern ohne wirklich freie Meinungsbildung und eingeschränktem Wahlwettbewerb.⁷ Parallel dazu wuchs auch die Zahl der Länder, die Referenden durchführen, besonders in den 1990er-Jahren wuchs die Zahl der Unabhängigkeitsreferenden,⁸ aber auch der Referenden in nicht-demokratischen Regimes.⁹ Und weil nicht-demokratische Regimes begannen demokratische Institutionen zu imitieren, wurde die Grauzone zwischen Demokratie und Diktatur zum neuen Forschungsfeld.¹⁰

-
- 6 CRTA, *Izveštaj o posmatranju referenduma za promenu Ustava*, (Beograd: CRTA), 2022.
 - 7 A. Schedler (Hrsg.), *Electoral Authoritarianism: The Dynamics of Unfree Competition* (Boulder: Lynne Rienner), 2006.
 - 8 F. Mendez und M. Germann, Contested Sovereignty: Mapping Referendums on Sovereignty over Time and Space, *British Journal of Political Science* 48, 2016, S. 141.
 - 9 M. Qvortrup, The Rise of Referendums. Demystifying Direct Democracy, *Journal of Democracy* 28(3), 2017, S. 141. Volle Demokratien halten aber global weiterhin viel mehr Referenden ab als autoritäre oder hybride Regimes D. Altman, *Direct Democracy Worldwide* (Cambridge: Cambridge University Press), 2010, S. 68.
 - 10 L. Diamond, Thinking about hybrid regimes, *Journal of Democracy* 13(2), 2002, S. 21; S. Levitsky und L. A. Way, The Rise of Competitive Authoritarianism, *Journal of Democracy*, 13(2), 2002, S. 51.

Besonders auffällig für Irregularitäten dürften Unabhängigkeitsreferenden und Verfassungsreferenden in Transformationsländern sein.¹¹ Die vermutete Anfälligkeit resultiert einerseits daraus, dass es bei Unabhängigkeitsreferenden und Verfassungsreferenden um sehr grundlegende, womöglich um Alles-oder-nichts-Entscheidungen geht, und wegen der Wichtigkeit derselben politische Akteure versucht sein könnten, alle in ihrer Gewalt stehenden Mittel einzusetzen, unbeachtet ob sie der Demokratie Schaden zufügen könnten oder nicht. Andererseits finden solche Referenden in Perioden des politischen Umbruchs statt. Sprich, in Ländern oder Regionen ohne gefestigte Institutionen, womöglich in Staaten oder Gebieten in denen die Aktivbürgerschaft (die Benennung der Bürgerinnen und Bürger, denen aktive politische Rechte zustehen) nicht klar geregelt ist, und oftmals im Kontext von starker sozialer Mobilisierung, Polarisierung oder im Nachgang zu Krieg. Mir ist keine systematische Vergleichsstudie zum Auftreten von Unregelmäßigkeiten in Volksabstimmungen bekannt. Es gibt aber eine Kluft zwischen beidseitig vereinbarten Referenden – die Konfliktparteien würden sich darauf ohne Garantien auf eine transparente, faire und akribisch überwachte Durchführung wohl kaum einigen – und unilateral angesagten Verfassungs- oder Unabhängigkeitsreferenden.¹² Ein indirekter Indikator dafür ist auch die unterschiedliche Konflikthaftigkeit von beidseitig vereinbarten oder unilateralen Unabhängigkeitsreferenden. Aus Letzteren sind oftmals Manipulationen bekannt, bei denen das Augenmerk offenbar nicht auf Discretion gerichtet war. Demgegenüber ist bei beidseitig akzeptierten und verhandelten Unabhängigkeitsreferenden die Integrität des Verfahrens zentral.

Für die wissenschaftliche Forschung, die sich mit der Integrität von Wahlen beschäftigt, die „Wahlforensik“, ist die kaum überblickbare Zahl der Möglichkeiten wie Wahlen und Referenden irregulär und illegitim verfälscht werden können, eine Herausforderung. Untersuchungen der Integrität von Wahlen und Abstimmungen sind zwangsweise stets auf einzelne Abläufe und einzelne Merkmale (Variablen) beschränkt. Die Vielzahl erschliesst sich aus den sieben Schritten, die Andreas Schedler als erforderlich ansieht, damit eine Wahl regulär abläuft. Ich übertrage hier sechs der sieben Schritte sinngemäss auf Abstimmungen.¹³

11 Mendez und Germann, *Contested Sovereignty: Mapping Referendums on Sovereignty over Time and Space* (Fn 8).

12 L. T. Farley, *Plebiscites and Sovereignty: the Crisis of Political Legitimacy* (London: Westview), 1986.

13 A. Schedler, *The Menu of Manipulation*, *Journal of Democracy* 13(2), 2002, S. 36.

1. Welche Vorlagen kommen zur Abstimmung, und in welcher Kombination? Wer formuliert die Abstimmungsvorlagen und die Abstimmungsfragen, und wie wird der Wahlzettel gestaltet?
2. Haben alle Seiten die nötigen politischen Rechte (z.B. Versammlungsfreiheit) und gleiche Chancen, einen Abstimmungskampf zu führen? Haben alle Seiten gleichberechtigt Zugang zu (öffentlichen) finanziellen Ressourcen, und können sie ihre Position in den Medien darstellen?
3. Haben alle Bürgerinnen und Bürger gleiche Rechte zur Stimmabgabe?
4. Wird auf die Stimmberechtigten Druck oder Zwang ausgeübt? Werden Stimmen gekauft?
5. Zunächst, werden alle Stimmen gezählt, und entspricht das Resultat den abgegebenen Stimmen? Wird das Resultat korrekt übermittelt? Weiter stellt sich aber auch eine ganze Reihe von Fragen bezüglich Entscheidungsprozeduren und -hürden in Referenden: Wie ist etwa in Souveränitätsreferenden mit sich widersprechenden Mehrheiten (betroffene Region versus Gesamtstaat) umzugehen? Wie werden Variantenabstimmungen mit zyklischen Ergebnissen entschieden? Lädt ein Beteiligungsquorum zur strategischen Stimmenthaltung ein, oder wird ein Beteiligungsquorum durch nicht-aktualisierte Wählerregister verfälscht?
6. Wird der Abstimmungsentscheid, auch wenn er gegen den Willen der repräsentativen Institutionen gefällt wurde, von Letzteren umgesetzt – insbesondere bei verbindlichen Volksentscheiden?¹⁴

Schedler's erstes Kriterium wird in dieser verkürzten Liste nicht aufgeführt, weil es kann nicht so einfach auf Abstimmungen übertragen werden. Es lautet, dass eine reguläre Wahldemokratie nur dann gegeben ist, wenn das Volk das Recht hat, mittels Wahlen alle Bereiche der Politik zu beeinflussen, sprich, dass keine Politikbereiche oder Posten der Kontrolle durch Wahlen entzogen sind. Dieser Aspekt ist in der Anwendung auf Abstimmungen umstritten. Übertragen auf Volksabstimmungen ergäbe sich daraus das radikal-direktdemokratische Recht, jederzeit Referenden und Volksinitiativen in allen Politikbereichen durchführen zu können, also auch zu Staatsfinanzen, zur Beschneidung von Minderheitenrechten, oder bezüglich internationaler Abkommen, die in manchen Ländern, die Volksabstimmungen praktizieren, von der direkten Demokratie ausgenommen sind.

14 E. R. Gerber, A. Lupia, M.D. McCubbins, und D.R. Kiewiet, *Stealing the Initiative. How State Government Responds to Direct Democracy* (Upper Saddle River (NJ): Prentice-Hall), 2001. Zur Schweizer Praxis: D. Bochsler, Das Schweizer Initiativrecht ist eine Fehlkonstruktion, *NZZ am Sonntag*, 2.03.2014.

Die lange Liste der Anforderungen zeigt primär auf, wie isoliert einzelne statistische Tests der Wahlforensik sind. Dies insbesondere angesichts der Erkenntnis aus der Wahlforschung, dass sich seit den 1990er-Jahren die gravierendsten Probleme mit Wahlen vom Wahltag (Schritt 5) in die Vorwahlperiode verschieben, und damit weniger gut beobachtbar sind.¹⁵ Für Abstimmungen dürften die Trends ähnlich verlaufen. Im verbleibenden Teil dieses Aufsatzes stelle ich einige der verwendeten statistischen Tests vor.

C. Gauss und die Wahlstatistik: Beispielhaftes zur Methode

Was im russischen Zarenreich die *potemkinsche Fassade*, sind in Putins Russland Wahlen. Und im Verlauf seiner vierten und gemäss Verfassung letzten Amtszeit hielt das Regime im Jahr 2020 ein Verfassungsreferendum ab. Dieses hatte den vielbeachteten Nebeneffekt, dass die Zählung der präsidentiellen Amtszeiten auf Null gestellt wurde. So kann Putin sein Amt bis ins Jahr 2036 verlängern – immer unter dem Vorbehalt, dass er sich in den Wahlen 2024 und 2030 wiederum zum Sieger erklärt. Das Forschungsteam Kobak, Shpilkin & Pshenichnikov hat die Resultate des Verfassungsreferendums aus dem Jahr 2020 genauer unter die Lupe genommen.¹⁶ Sie weisen auf ähnliche Auffälligkeiten hin, wie sie auch bei Wahlen immer wieder auftreten. Demnach häuften sich Zählkreise,¹⁷ die eine Zustimmung von 85% registrierten. Solange eine lokale Wahlbehörde rapportiert, dass von 1020 Stimmen genau 867 ein „Ja“ einlegten, sieht dies unverdächtig aus, und auch der Zustimmungswert von punktgenau 85.0% ist plausibel. Auffallend war die landesweite Häufung von Zählkreisen mit sehr runden Ja-Anteilen (85.0%, 90.0%, etc.). In der Kleinstadt Klinty rapportierten die meisten Zählkreise eine Stimmbeteiligung von haargenau 90.0% oder 91.0%, während in der Stadt Nalchik die meisten Zählkreise eine 80.0%-ige Wahlbeteiligung, mit 90.0%-iger Zustimmung berichteten. So unauffällig die absoluten Zahlen aussehen mögen, in Prozentwerte umgerechnet, ist

15 P. Norris, *Why Electoral Integrity Matters* (Cambridge: Cambridge University Press), 2014, S. 84.

16 D. Kobak, S. Shpilkin, und M. S. Pshenichnikov, *Suspect peaks in Russia's "referendum" results*, *Significance* 17, 2020, S. 8.

17 Zählkreise sind die kleinste Einheit, für die Abstimmungsergebnisse ausgezählt werden, oftmals das Einzugsgebiet eines Stimmlokals; in der Regel entsprechen sie einer Wahlurne. Insb. in urbanen Gebieten können Stimmlokale auch mehrere Wahlurnen und damit mehrere Zählkreise aufweisen.

die Verteilung viel zu uniform, als dass sie nicht handgefertigt wäre. Möglich ist etwa, dass eine übergeordnete Instanz den lokalen Wahlleitern Prozentquoten für die Zustimmung vorgegeben hat, und viel zu viele von ihnen diese Zielwerte punktgenau erreicht haben, oder aber dass eine übergeordnete Instanz die Wahlresultate im Nachhinein verändert hat.¹⁸

Diese Vermutungen erstaunen umso mehr, wenn man sich vor Augen führt, dass die Amateurhaftigkeit solcher Wahlergebnisse der Wahlbehörde schon lange bekannt sein muss. Nach der Duma-Wahl 2011 erlebte Russland die grösste Protestwelle seit den 1990er-Jahren; die Opposition sprach von einer gestohlenen Wahl. Ein Motiv der Proteste zeigte den deutschen Mathematiker Carl Friedrich Gauss, den Vater der statistischen Normalverteilung. Illustriert war das Poster mit einer grafischen Auswertung der Wähleranteile der Partei des Präsidenten, Einiges Russland, aus Zehntausenden von Wahllokalen quer durch Russland. Wie Stacheln ragen aus der Gauss'schen Verteilung in Zehn-Prozent-Schritten Spitzen heraus, Abweichungen, und zwar nicht zufällig, sondern haarscharf bei 50.0%, 60.0%, 70.0%, 80.0%, 90.0% und 100.0%. Doch so flächendeckend wie im Referendum von 2020, mit „Anomalien“ von geschätzt 3670 Wahllokalen, wurde das Fälschungsschema noch nie angewandt.¹⁹

I. Reflexionen zur Methodologie

Das dargelegte Verfahren ist in mancher Hinsicht bemerkenswert, und eignet sich vorzüglich zur Diskussion der Methodologie der Wahlforensik.

Zunächst ist es nicht zufällig, dass das Forscherteam um Kobak die Verfassungsabstimmung mit einem Verfahren analysiert, welches sie zuvor für die Überprüfung von Wahlresultaten verwendet und perfektioniert hat. Die Literatur zur Wahlforensik – der Begriff besagt es – beschäftigt sich zunächst mit der Integrität von Wahlen.²⁰ Ich kenne aus methodologischer

18 D. Kobak, S. Shpilkin, und M. S. Pshenichnikov, Integer Percentages as Electoral Falsification Fingerprints, *The Annals of Applied Statistics* 10(1), 2016, S. 54.

19 Kobak et al., Suspect peaks in Russia's "referendum" results (Fn 16).

20 R. M. Alvarez, T. E. Hall, und S. D. Hyde, Introduction: Studying Election Fraud, in R. Michael Alvarez, Thad E. Hall, und Susan D. Hyde (Hrsg.), *Election Fraud. Detecting and Detering Electoral Manipulation* (Washington (DC): Brookings Institution), 2008, S. 1; C. Breunig und A. Goerres, Searching for electoral irregularities in an established democracy. Applying Benford's Law tests to Bundestag elections in Unified Germany, *Electoral Studies* 30(3), 2011, S. 534; J. Deckert, M. Myagkov, und P. Ordeshook, Benford's Law and the detection of election fraud, *Political Analysis* 19(3), 2011, S. 245; W. R. J. Mebane, Election forensics:

Sicht höchstens periphere Gründe, warum die gleichen Methoden nicht auch für die Analyse von Volksabstimmungen geeignet sein sollten. Zu den substanziellen Unterschieden zwischen Wahlen und Volksabstimmungen gehört erstens, dass in Wahlen viel mehr Variablen zu untersuchen sind (etwa die Stimmzahlen für alle Parteien, die Wahlbeteiligung, sowie leere und ungültige Wahlzettel), die voneinander abhängig sind. Manche Länder oder Bundesstaaten bringen gleichzeitig mehrere, inhaltlich voneinander unabhängige direktdemokratische Vorlagen zur Abstimmung, und Unabhängigkeit der Sachfragen mag für manche Quervergleiche vorteilhaft sein. Zweitens spielen bei Wahlen die Wahlkreise eine gewichtige Rolle, so dass manche Abweichungen von normalen Verteilungen womöglich auf Effekte der strategisch bestimmten Wahlkreisgrenzen zurückzuführen sind. Doch während für Referenden in aller Regel solche administrativen Grenzen keinen Einfluss auf das Resultat haben, und direkte strategische Effekte daher wenig plausibel sind, werden Referenden meistens von den gleichen Wahlbehörden und in den gleichen administrativen Zählkreisen durchgeführt und ausgezählt. Somit ist nicht ausgeschlossen, dass sich potenzielle administrativ-strategische Artefakte aus Wahlen auch auf Abstimmungen übertragen. In Kürze: es gibt kaum Gründe, die Abstimmungsforensik neu zu erfinden – und in der Tat werden alle in diesem Aufsatz beschriebenen Verfahren sowohl für Wahlen als auch für Abstimmungen verwendet.

Weiter lassen sich anhand des dargestellten Verfahrens einige Gütekriterien von wahlforensischen Verfahren darstellen.

Erstens erscheint das Vorgehen geeignet, um nicht natürlich auftretende Abweichungen zu erkennen, die darauf hindeuten, dass das rapportierte Abstimmungs- oder Wahlergebnis in einem Teil der Zählkreise handgefertigt wurde. Es gibt gemäss meiner Kenntnis keine alternative Erklärung für solche Muster, sprich keinen Grund davon auszugehen, dass solche Verteilungen natürlich auftreten. Dies ist eines der beiden wichtigsten Kriterien zur

the second-digit Benford's law test and recent American presidential elections, in Michael Alvarez, Thad E. Hall, und Susan D. Hyde (Hrsg.), *Election Fraud: Detecting and Deterring Electoral Manipulation* (Washington (DC): Brookings Institution), 2008, S. 162; M. Myagkov, P. Ordeshook, und D. Shakin, *The Forensics of Election Fraud. Russia and Ukraine*, Cambridge: Cambridge University Press, 2009; F. M. Sjöberg, Autocratic adaptation: The strategic use of transparency and the persistence of election fraud, *Electoral Studies* 33, 2014, S. 233; P. Norris, *Why Electoral Integrity Matters*; J. Medzihorsky, Election Fraud: A Latent Class Framework for Digit-Based Tests, *Political Analysis*, 2015; D. Bochsler, J. Medzihorsky, C. Schürmann und P. Stark, *Report on the Identification of Electoral Irregularities by Statistical Methods* (Strasbourg: Council of Europe, European Commission for Democracy Through Law (Venice Commission)), 2018.

Beurteilung der Messgüte statistischer Methoden zur Erkennung potenzieller Irregularitäten in Wahlen und Abstimmungen. Denn Methoden, die ganz natürliche politische Phänomene als Manipulation missdeuten, wären für die Wahlforensik wenig geeignet (hierzu der Fall Boliviens, Abschnitt E). Darüber hinaus gibt es eine wichtige Einschränkung, die universell gilt. Selbst wenn bei einem Abstimmungsresultat ohne jeden begründeten Zweifel von einem menschlichen Artefakt ausgegangen werden kann, lässt sich daraus noch nicht auf Wahlfälschung schliessen; sie könnten auch einfach menschlicher Irrtum, oder das Ergebnis administrativer Nachlässigkeit (etwa einer Schätzung anstelle einer Auszählung des Resultates) sein (siehe Texas 1860, Abschnitt E). Universell gilt ferner, dass alle statistischen Methoden – die eingeführte wie andere – Annahmen treffen über eine 'natürliche' Verteilung der Wahl- und Abstimmungsresultate. Solche können aus früheren Urnengängen oder allgemeinen mathematischen Gesetzen hergeleitet werden. Darauf aufbauend werden deutliche und häufige Abweichungen identifiziert, und die Plausibilität, und mögliche alternative Erklärungen der Abweichungen, diskutiert.²¹

Ein zweites Kriterium der Messgüte betrifft, wie gut die Methode Irregularitäten erkennt. Hier wäre anzumerken, dass die Methode nur eine sehr spezifische Form von Manipulationen erkennt, welche die Auszählung oder Übermittlung der Stimmen betrifft, und auch hier nur ein sehr spezifisches (mutmasslich sehr seltenes) Vorgehen der Wahlbehörden offenlegt. Die allermeisten Unregelmässigkeiten – insbesondere diejenigen vor der Auszählung, aber auch die meisten möglichen Probleme bei der Auszählung (siehe Schedler's Liste in Abschnitt B) – bleiben unerkannt. Daher können weder negative Befunde dieses Prüfverfahren hier noch jegliche andere Verfahren dazu dienen, eine Abstimmung als 'sauber' (manipulationsfrei) zu apostrophieren. Sie können bloss einen allfälligen Verdacht ausräumen, dass eine ganz spezifische Manipulation systematisch vorgenommen wurde.²²

Drittens basiert auch jede bestens ausgeklügelte und valide Methode immer auf Wahrscheinlichkeiten. Die identifizierten Abweichungen können auch rein zufällig erfolgen. Die Statistik gibt eine Wahrscheinlichkeit an, aufgrund der eine Verteilung der Erwartung folgt, oder davon abweicht (ob aufgrund menschlicher Manipulation oder natürlicher Prozesse sei dahin-

21 D. Baum, Pinpointing Apparent Fraud in the 1861 Texas Secession Referendum, *Journal of Interdisciplinary History* 32(2), 1991, S. 201.

22 Bochsler et al., Report on the Identification of Electoral Irregularities by Statistical Methods (Fn 20).

gestellt). Sprich, die identifizierten Abweichungen können immer auch zufällig erfolgt sein. Analysieren wir eine grosse Zahl von Abstimmungen, oder innerhalb einer Abstimmung eine grosse Zahl von Merkmalen (etwa Ja-/Nein-Stimmen oder gar Variantenabstimmungen, leere und ungültige Stimmzettel, Wahlbeteiligung) steigt die Wahrscheinlichkeit, dass eine rein zufällig vorkommende Verteilung bei einer der mehreren zu analysierenden Variablen einem der Muster entspricht, das unsere Methode als auffällig interpretiert. Leemann und Bochslers passen daher die Testkriterien so an, dass bei einer grösseren Zahl parallel durchgeführter statistischer Tests auch das Mass der (vermeintlichen) Abweichungen von der Verteilung höher angesetzt werden muss.²³

Und viertens arbeiten auch in manchen fehlhaften Wahlbehörden Fachleute. Sie können sich über die Testmethoden informieren, und ihre Resultate dagegen immunisieren (siehe Schlussfolgerungen).

D. Verteilungen und Zahlengesetze

In der Umfrageforschung leistete der Soziologe Andreas Diekmann Pionierarbeit, um aus Zahlen herauslesen, ob sie authentisch sind oder nicht.²⁴ Er stützte sich dabei auf ein Zahlengesetz, das auf sehr viele natürliche Zahlen Anwendung findet, wie sie in Umfragen häufig erfragt werden. Das Gesetz kann auch auf (positive) Zahlen aus dem Alltag angewendet werden, wie den Saldo auf dem privaten Bankkonto, oder auf die Distanz zwischen Wohn- zum Arbeitsort, und zwar unabhängig von der Währung, in der das Bankkonto geführt bzw. dem Mass (Meter, Ellen, Inches), in welchem gerechnet wird. Wenn eine grosse Zahl solcher Zahlen vorhanden ist, sollten diese einem natürlichen Gesetz zur Verteilung der ersten Ziffern folgen. Gemäss Benford's Gesetz beginnen all diese Zahlen etwas häufiger mit der Ziffer 1 als mit darauffolgenden Ziffern. Am seltensten steht eine 9 an erster Stelle. Daraus lässt sich die erwartete Verteilung der ersten Ziffer berechnen; und natürlich ebenso für die zweite, dritte und weitere Zahlenstellen. Für diese nähert sich das Gesetz einer gleichförmigen Verteilung an. Davon ausgenommen sind Zahlen mit einer natürlichen Obergrenze, also beispielsweise Telefonnummern oder das Lebensalter.

23 L. Leemann und D. Bochslers, A Systematic Approach to Study Electoral Fraud, *Electoral Studies* 35, 2014, S. 33.

24 A. Diekmann, Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, *Journal of Applied Statistics* 34, 2007, S. 321.

Dieses Gesetz wendete Diekmann auf Umfragedaten an, im Wissen darum, dass es unter den Mitarbeitenden von Umfrageinstituten ehrliche gibt, und andere, die einzelne Zahlen im Fragebogen selber erfinden. Sobald Menschen Zahlen frei erfinden, verfehlen sie die Benford-Verteilung; in grösseren Zahlenmengen lässt sich diese Abweichung messen.

Die Benford-Methode erfreut sich in der Wahlforensik grosser Beliebtheit. In ihrer Anwendung auf mehrere Wahlen und Abstimmungen auf dem amerikanischen Kontinent stossen Pericchi und Torres²⁵ sowie Jiménez²⁶ im Venezolaner Referendum von 2004 zur Abberufung des Präsidenten auf eine auffällig gleichförmige Verteilung von Abstimmungsergebnissen. Die Stimmabgabe erfolgte teils manuell, teils elektronisch. Und während die auf Papier abgegebenen Stimmen regulär zu sein schienen (immer in Bezug auf das Benford-Gesetz), war die jeweils zweite Stelle der sich auf die elektronische Stimmabgabe beziehenden Zählkreisresultate auffällig gleichmässig. Dieser Umstand wäre gemäss Wahrscheinlichkeitsrechnung bei einer regulären Zählung kaum möglich. Gemäss Benford müssten eigentlich die Null und die Eins etwas häufiger vorkommen, und die Neun am wenigsten oft.

Das Venezolaner Referendum wurde in der Forschung weit beachtet, der Benford-Test wurde aber auch anderswo angewandt, beispielsweise bei russischen Wahlen, und ist inzwischen weit bekannt. Offenbar auch bei den russischen Wahlbehörden. Kalinin und Mebane haben festgestellt, dass in den russischen Parlamentswahlen 2016 die letzte Ziffer haargenau dem theoretischen Erwartungswert entsprach.²⁷ Wiederum fiel das Resultat auf: statistische Verteilungen folgen nie punktgenau der theoretischen Verteilung, sondern weisen geringe Abweichungen auf. Die Verteilung liefert einen Hinweis darauf, dass eine zentrale Stelle die Ergebnisse aus den knapp 97'000 Zählkreisen genau unter die Lupe genommen hat, um statistische Auffälligkeiten, die auf Fälschungen hinweisen, zu entfernen, ihre Arbeit aber zu gründlich gemacht hat, ja gar auffallend gründlich.

25 L. Pericchi und D. Torres, "Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela," *Statistical Science* 26(4), 2011, S. 502.

26 R. Jiménez, "Forensic Analysis of the Venezuelan Recall Referendum," *Statistical Science* 26(4), 2011, S. 564.

27 K. Kalinin und W. R. J. Mebane, *When the Russians fake their election results, they may be giving us the statistical finger*, <https://www.washingtonpost.com/news/monkey-cage/wp/2017/01/11/when-the-russians-fake-their-election-results-they-may-be-giving-us-the-statistical-finger/> (Zugriff 10.6.2022).

E. Analyse der politischen Geografie

Für die meisten Formen des Wahlbetrugs gibt es keine statistischen Nachweise. So kann etwa grossflächiger Stimmenkauf nicht mittels statistischer Methoden nachgewiesen werden. Wir wissen zwar, dass marginalisierte, bildungsferne Bevölkerungsteile auf Geschenke oder finanzielle Tauschgeschäfte nicht verzichten können, und so leichte Beute für Stimmenkäufer sind. Manche Wahlbeobachterinnen und Wahlbeobachter haben ein gutes Auge dafür: sie stellen etwa Leute fest, die auffällig unauffällig vor Wahllokalen patrouillieren. Es wäre ein Leichtes eine statistische Methode zu entwickeln, die Wählerströme aus Armutsvierteln identifiziert, und diesen den Stempel des Wählerkaufs aufdrückt. Damit würden wir aber Parteien, die sich redlich um eine Verbesserung der Lebensverhältnisse der Ärmsten bemühen, und dafür in Wahlen und Abstimmungen belohnt werden, zu Unrecht in Verruf bringen.

Die angesprochene Analyseverfahren – der Vergleich von sozioökonomischen Daten mit Abstimmungsergebnissen – beruht auf der politischen Geografie, und gehört damit zur zweiten Familie von wahlforensischen Analyseverfahren. Neben soziologischen Aspekten der politischen Geografie gehören hierzu auch (temporale) Vergleiche von Wahl- und Abstimmungsergebnissen. Stets wird das Stimm- oder Wahlverhalten in den Zählkreisen mittels Resultaten aus früheren Wahlen und Abstimmungen erklärt, mittels der soziologischen Struktur, oder aber es lassen sich mehrere Vorlagen (oder Teilfragen) vergleichen, über die gleichentags abgestimmt wird, oder gar mehrere Variablen (z.B. Beteiligung und Zustimmung) derselben Vorlage. In allen Fällen wird ein Modell geschätzt, das die territorialen Unterschiede in den Wahl- und Abstimmungsergebnissen zu erklären versucht, meistens die Unterschiede zwischen den Zählkreisen, manchmal aber auch auf höheren Aggregationsebenen (Gemeinden, Bezirke). Vermeintliche Auffälligkeiten in Wahl- und Abstimmungsergebnissen können sich als Folge natürlicher Phänomene entpuppen, etwa der Wahlgeografie.²⁸ Klimek und Koautoren weisen beispielsweise in einer nicht publizierten Studie (Preprint) zum türkischen Verfassungsreferendum von 2017 auf eine verdächtige Korrelation von uniformer Stimmabgabe mit hoher Wahlbeteiligung in

28 Bochslers et al., Report on the Identification of Electoral Irregularities by Statistical Methods (Fn 20).

einem Teil der Zählkreise hin.²⁹ Sie werten dies als Hinweis darauf, dass die lokale Wahlbehörde im Namen der Nicht-Stimmenden regierungstreue Wahlzettel eingeworfen hat. Doch territoriale Unterschiede – etwa in der Homogenität der Wahlresultate – brauchen nicht unbedingt das Resultat lokaler Manipulationen von Abstimmungsresultaten zu sein: es gibt auch einen allgemeinen soziologischen Trend zu stärkerer politischer Kohäsion in peripheren Regionen, kleineren Ortschaften, oder in Gebieten mit einer spezifischen sozioökonomischen Struktur oder einem hohen Anteil ethnischer Minderheiten.³⁰ Die Korrelation hoher Wahlbeteiligung mit der Stimmabgabe für eine bestimmte Partei oder eine der Optionen auf dem Referendums-Stimmzettel kann das Resultat einseitiger Mobilisierung sein. Korrelationen von Wahlbeteiligung und Parteistärke sind auch in alten Demokratien üblich.³¹

Mit diesem Problem sind alle Analyseverfahren behaftet, die versuchen aufgrund der Wahlgeografie plausiblere von weniger plausiblen Resultaten zu unterscheiden. Die Geografie selbst liefert eine Vielzahl von Erklärungen, und letztlich ist fraglich, was auf eine Fälschung hindeutet, und was auf natürliche oder strategische Effekte. Erik Herron betrachtet in seiner Analyse von Wahlen und eines Referendums in Aserbaidschan unter anderem die Wahlbeteiligung, die Häufung von Stimmen in Abwesenheit (mobile Wahlurnen oder Stimmabgabe in anderen Wahllokalen), und wertet besonders hohe Werte auf den beiden Variablen als mögliche Indikatoren für eine Wahlfälschung.³² Allerdings sind solche Indikatoren sehr zweifelhafte Beweise von Wahlbetrug. Es ist daher darauf hinzuweisen, dass Herron nicht von Wahlbetrug spricht, sondern diese Variablen nur indirekt verwendet, um potenziell problematische Stimmlokale zu erfassen.

29 P. Klimek, R. Jiménez, M. Hidalgo, A. Hinteregger und S. Thurner, Election forensic analysis of the Turkish Constitutional Referendum 2017, 2017, arXiv preprint.

30 S. Coleman, The Effect of Social Conformity on Collective Voting Behaviour, *Political Analysis* 12, 2004, S. 76. Bemerkenswert scheint, dass Klimek et al. (Fn 29) selber feststellen, dass das Muster in kleineren und peripheren Wahllokalen auftritt, aber schreiben hierzu, dass in kleineren Wahllokalen Manipulationen einfacher durchführbar sind. Damit orientieren sie sich sicherlich nicht am Prinzip 'in dubio pro reo'.

31 B. Grofman, G. Owen, und C. Collet, Rethinking the partisan effects of higher turnout: So what's the question?, *Public Choice* 99, 1999, S. 357.

32 E. S. Herron, Measuring Dissent in Electoral Authoritarian Societies: Lessons From Azerbaijan's 2008 Presidential Election and 2009 Referendum, *Comparative Political Studies* 44(11), 2011, S. 1557.

Im Bewusstsein dieses Problems verwenden andere Autorinnen und Autoren die Geografie der Wahlergebnisse nur als ein Hilfselement, das kombiniert mit weiteren Kriterien durchaus interessant wird, etwa im Fall des oben erwähnten Venezolaner Recall-Referendums. Unter anderem haben Taylor,³³ Febres Cordero und Márquez³⁴ die Resultate aus all jenen Wahllokalen genauer analysiert, in denen mehrere elektronische Wahlmaschinen zum Einsatz kamen. Die Stimmenden werden in einem Verfahren, das auf der Personen-Identifikationsnummer basiert, zufällig einer der mehreren Wahlmaschinen zugewiesen; daher sollten die Resultate der einzelnen Wahlmaschine nur im Zufallsbereich voneinander abweichen. Die Analysen kommen zu widersprüchlichen Resultaten, was auf das unterschiedliche Mass an methodologischer Stringenz der Studien zurückzuführen sein dürfte. Jiménez zeigt in einer akribischen Analyse beispielsweise auf, dass eine Studie fälschlicherweise geografisch benachbarte Wahlmaschinen miteinander verglichen haben, die aber nicht im gleichen Wahllokal platziert waren, womit diese Zufallszuteilung nicht zutrifft. Andere haben zu wenig komplexe statistische Modellierungen verwendet haben, oder er weist darauf hin, dass sich die Interpretation ändern kann, je nachdem ob nur die Ja- und Nein-Stimmen zwischen den Wahlmaschinen verglichen werden, oder ob auch die Enthaltungen in der Analyse berücksichtigt werden.³⁵

1861 führte Texas ein Referendum über eine mögliche Sezession durch. Dale Baum hat die Ergebnisse auf Ebene der Bezirke mit den Ergebnissen der vorhergehenden Präsidentschaftswahl von 1860 verglichen. Während diese Wahl insgesamt eine gute Voraussage ermöglicht, sowohl hinsichtlich der Zustimmung zum Referendum als auch bezüglich der Stimmbeteiligung, tanzten rund zwei Dutzend Bezirke aus der Reihe, sprich sie hatten grosse Residuen bei einer der beiden abhängigen Variablen (Zustimmung und Beteiligung), oder gar bei beiden. Baum untersuchte diese qualitativ genauer, um Hinweise zu finden, die die Ausreisser erklären könnten. In manchen Fällen handelte es sich um offensichtlichen menschlichen Irrtum – etwa bei der Übermittlung der Resultate. In anderen Fällen stiess er in der lokalen Presse auf Hinweise auf mögliche Unregelmässigkeiten. In dieser Kombination – statistische Abweichungen und Hinweise auf Betrug,

33 J. Taylor, *Too many ties? An empirical analysis of the Venezuelan recall referendum counts* (Stanford (CA): Stanford University), 2005.

34 M. M. Febres Cordero und B. Márquez, A Statistical Approach to Assess Referendum Results: the Venezuelan Recall Referendum 2004, *International Statistical Review* 74(3), 2006, S. 379.

35 Hierzu: Jiménez, Forensic Analysis of the Venezuelan Recall Referendum (Fn 26).

und zwar genau in denjenigen Gebieten, die durch die Statistik ausgewiesen wurden – schien letzterer naheliegend.

Ganz im Sinne der Texanischen Analyse verbinden Leemann und Bochsler geografische Merkmale mit systematischen Merkmalen, die mit Unregelmässigkeiten zusammenhängen könnten, in einer Untersuchung einer Volksabstimmung im Schweizer Kanton Bern.³⁶ Dort wurde über eine ökologische Reform der Motorfahrzeugsteuer abgestimmt, mit sehr knappem Ausgang, worauf die oberste Justizbehörde eine Nachzählung anordnete. Doch als diese beschlossen war, hatten 30 Gemeinden (von 383 Gemeinden) die Wahlzettel bereits vernichtet, entgegen der Weisung der Wahlbehörde diese angesichts des laufenden Justizverfahrens aufzubewahren. Damit war die Nachzählung nicht mehr möglich. Gegenüber den Untersuchungsbehörden erklärte jede einzelne der fehlbaren Gemeinden, wie es zur Vernichtung der Wahlzettel kam. In den meisten Fällen machten die Gemeinden das Reinigungspersonal, Auszubildende oder andere Subalterne dafür verantwortlich. Die dreissigste Gemeinde fand die Wahlzettel wieder – zu einem Zeitpunkt als die Nachzählung hinfällig geworden war. Leemann und Bochsler gingen der Frage nach, ob die zunächst rapportierten Resultate Hinweise auf Unregelmässigkeiten in den 30 fehlbaren Gemeinden (identisch mit 30 Zählkreisen) lieferten.

Einerseits identifizieren sie eine Reihe von Zählkreisen (hier: Gemeinden), für die externe, also nicht aus der Statistik gewonnene, Hinweise auf seltsame Abläufe nach dem Wahltag vorlagen. Andererseits erlaubt es ihnen die Zahlenlage, die Resultate der jeweiligen Zählkreise nicht nur aufgrund historischer Vergleichswahlen oder -abstimmungen vorherzusagen, sondern vielmehr schätzen sie auch Modelle, die die verschiedenen Variablen, die aus dem entsprechenden Wahltag resultieren (hier: mehrere Abstimmungsfragen) zueinander in Verbindung stellt. Zunächst wenden sie das Verfahren in einer Placebo-Analyse auf frühere Abstimmungen an, um die Validität der Methode zu überprüfen, also insbesondere um mögliche Typ I-Fehler (falsche Positive) zu erkennen. Allen Methoden der Wahl- und Abstimmungsforensik haftet der Verdacht an, dass vermeintlich auffällige Werte ebenso gut das Ergebnis natürlicher Phänomene oder strategischen Akteurverhaltens, inkl. der strategischen Festlegung von Wahlkreisen und Zählkreisen, sein könnten.

Dabei kommen ihnen drei spezifische Gegebenheiten zu Hilfe. Erstens handelt sich um ein weltweit rares Beispiel von Variantenabstimmungen, also einer Abstimmung, bei welcher die Stimmberechtigten zwischen mehr

36 Leemann und Bochsler, A Systematic Approach to Study Electoral Fraud (Fn 23).

als nur einer Zustimmung oder Ablehnung entscheiden können.³⁷ Im fraglichen Fall konnten die Stimmberechtigten über zwei Versionen einer Gesetzesrevision befinden, was – zusammen mit einer Eventualfrage über die Präferenzen im Falle einer doppelten Annahme beider Gesetzesvarianten – zu drei Teilfragen auf dem Stimmzettel führte. Die Teilfragen stehen inhaltlich in einem Bezug zueinander, was sich auch in Korrelationen der Resultate der Teilaspekte der Abstimmung niederschlägt. Andernorts treten solche Korrelationen auch auf, sei es zwischen der Zustimmungsrate und der Stimmbeteiligung (wie im texanischen Beispiel), oder zwischen mehreren Sachvorlagen, und ggf. Wahlen, die am gleichen Tag zur Abstimmung kommen. Zweitens hatten sie aufgrund der 30 Gemeinden eine klar definierte, und zudem tatsächlich externe (nicht vom Abstimmungsergebnis hergeleitete) Verdachtsgruppe. Das mag aussergewöhnlich anmuten, aber es gibt in anderen Kontexten gleichwertige Indikatoren, etwa Wahllokale oder Gemeinden mit nicht-überparteilichen Wahlbehörden, Zählkreise in denen die Wahl angefochten wurde, örtliche Diskrepanzen zwischen Exit-Polls und echten Resultaten,³⁸ oder Zählkreise mit einer grossen zeitlichen Verzögerung bei der Auszählung.³⁹ Selbstverständlich sollte die Methode, die im zweiten Schritt zur Messung von Auffälligkeiten verwendet wird, nicht abhängig sein von den Merkmalen, die im ersten Schritt verwendet wurden um die Wahllokale mit einer erhöhten Anfälligkeit auf Manipulationen zu identifizieren. Es wäre also unzulässig, im ersten Schritt Abstimmungslokale zu denen Wahlbeschwerden eingegangen sind oder Hochburgen der Regierungspartei als mögliche Quellen von Unregelmässigkeiten zu identifizieren, falls diese in irgendeiner Art den im zweiten Schritt betrachteten statistischen Test beeinflussen könnten,⁴⁰ oder falls Abstimmungsbeschwerden aufgrund von statistischen Auffälligkeiten erhoben wurden. Drittens bietet die häufige Praxis von Volksabstimmungen in der

37 D. Bochsler, The Marquis de Condorcet goes to Bern, *Public Choice* 144(1-2), 2010, S. 119.

38 R. Prado und B. Sansó, The 2004 Venezuelan Presidential Recall Referendum: Discrepancies Between Two Exit Polls and Official Results, *Statistical Science* 26(4), 2011, S. 517.

39 Der Zeitpunkt der Berichterstattung des vorläufigen Wahlergebnisses stand auch im Zentrum der Kontroverse um die vermeintlich gefälschte Präsidentschaftswahl in Bolivien, 2019. Zur Identifikationsstrategie basierend auf dem Zeitpunkt der Resultatsübermittlung, siehe N. Idrobo, D. Kronick und F. Rodriguez, Do shifts in late-counted votes signal fraud? Evidence from Bolivia, *Journal of Politics* 84(4), 2022, S. 2202.

40 Beispielhaft: Jiménez, Forensic Analysis of the Venezuelan Recall Referendum (Fn 26).

Schweiz, und von Variantenabstimmungen im Kanton Bern, genügend Vergleichsdaten, einerseits zwecks Vorhersagung des Wahlergebnisses nach Zählkreisen und zur Identifizierung von Abweichungen, andererseits für die Placebo-Tests.

Resultate zeigen, dass in den Gemeinden, die ihre Wahlzettel verloren hatten, die Stichfrage weniger häufig offengelassen wurde. Die Stichfrage ist eine Besonderheit der Stimmtzettel, die in der Schweiz bei Variantenabstimmungen verwendet werden. Viele der Stimmenden lassen die Stichfrage offen, obwohl sie durchaus wichtig sein kann, im untersuchten Referendum gab sie gar den Ausschlag. In mehreren Gemeinden, die ihre Wahlzettel verloren hatten, war die Anzahl leerer Stichfragen klein, und zwar so klein, dass dies bei einer genaueren Analyse kaum erklärbar ist. Doch ob diese Abweichung tatsächlich statistisch signifikant war (also nicht bloss zufällig) ist im spezifischen Fall eine Interpretationsfrage, die mit dem Untersuchungsdesign zusammenhängt, und der Frage welcher statistische Test angewandt werden muss. Bei konservativer Lesart sind die Abweichungen zu wenig systematisch als dass sie erhärtete Schlüsse zulassen würden.

Der Ausgang direktdemokratischer Abstimmungen kann sowohl rechts als auch links in der Regressionsgleichung fingiert werden. Im bis dato womöglich wichtigsten und folgenschwersten Anwendungsfall der Wahlforensik, den bolivianischen Präsidentschaftswahlen 2019, stand das Verfassungsreferendum auf der rechten Seite der Gleichung, oder zu Deutsch, die Resultate des Referendums von 2016 dienten dazu, die Betrugsvorwürfe in den Präsidentschaftswahlen drei Jahre darauf zu entkräften. Die Organisation Amerikanischer Staaten (OAS), die die Wahl 2019 beobachtete, wies in ihrem Bericht auf statistische Auffälligkeiten im Wahlergebnis hin.⁴¹ Dies war ein innovativer Schritt – statistische wahlforensische Untersuchungen fließen sonst nicht in die Berichte von Wahlbeobachtungsmissionen ein.

Doch von vorne. Im Oktober 2019 stand der bolivianische Staatspräsident Evo Morales im ersten Wahlgang zur Wiederwahl. Die Möglichkeit einer dritten Amtszeit hatte er sich mit dem Verfassungsreferendum im Vorjahr eröffnet. Gemäss der zuerst eintreffenden provisorischen Teilergebnisse verpasste der linke Amtsinhaber Morales das notwendige Mehr von 40% für die Wahl im ersten Wahlgang. Im Verlauf der Wahlnacht trafen aber neue Resultate ein, und wendeten das Blatt allmählich zu Gunsten von Morales. Schliesslich unterbrach die Wahlkommission die Berichterstat-

41 Organization of American States *Electoral integrity analysis. General Elections in the Plurinational State of Bolivia. Preliminary Findings*, 2019, OAS.

tung über die Wahlergebnisse, wie sie es immer macht während Pressekonferenzen, doch dieses Mal gab es eine Pause von 23 Stunden bis die Resultate aller verbleibenden Wahllokale publiziert wurden, und nun stand Morales als Sieger fest. Die OAS nahm diesen Zeitsprung statistisch unter die Lupe. Den Sprung in den Stimmzahlen zum Zeitpunkt der Unterbrechung führte die Organisation Amerikanischer Staaten als Beweis für gross angelegten Wahlbetrug an. Es kam zu Unruhen; Polizei und Militär verweigerten Morales den Gehorsam, der Präsident wurde gestürzt.

Die Turbulenzen vom Oktober 2019 sorgten in der Weltöffentlichkeit für grosse Beachtung, der Bericht der Wahlbeobachter wurde sowohl in Bolivien als auch international als Beweis für den Wahlbetrug hochgehalten. Die darin verwendete Methode wurde meines Wissens noch nie zum Aufspüren von Wahlbetrug angewandt. Sie ist dafür auch nicht geeignet.⁴² Eine Analyse des Autorenteam's Idrobo, Kronick und Rodriguez, deutet darauf hin, dass der angebliche Beweis der Wahlfälschung einer genaueren Betrachtung nicht standhält. Das Gegenargument ist einfach und einleuchtend: weltweit treffen Wahl- und Abstimmungsresultate nach der Schliessung der Stimmlöcher immer gestaffelt bei der zentralen Wahlkommission ein. Das hängt mit landesspezifischen Faktoren zusammen, etwa der Grösse der Zählkreise oder der unterschiedlichen Komplexität des Zählprozesses, etwa wegen der Anzahl Vorlagen. Bei Fehlen elektronischer Übertragungsmöglichkeiten kommen die unterschiedlich langen Anfahrtswege hinzu. Wo auch immer gewählt oder abgestimmt wird, sind damit die zuerst eintreffenden Resultate (ohne Gewichtung oder Entzerrung) keineswegs repräsentativ. In Bolivien treffen zuerst Resultate aus den Zentren ein, und später die Resultate aus peripheren Gebieten, wo viele Indigene wohnen, also der Hausmacht des ersten indigenen Präsidenten des Landes, Evo Morales. Indiro's Forschungsteam verwendete das Resultat des thematisch ähnlich gelagerten Verfassungsreferendums von 2016 (dritte Amtszeit von Morales) als Vergleichsbasis für die Präsidentschaftswahlen, und wies nach, dass die politische Geografie der Präsidentschaftswahl 2019 ziemlich genau dem

42 Die Schätzung basiert auf einem Regressions-Diskontinuitätsdesign; mit der ordinalen Reihenfolge der eintreffenden Resultate als Erklärungsvariable. Die ordinale Reihenfolge vor der Unterbrechung der Berichterstattung dürfte eine Funktion der Geographie und der Grösse der Zählkreise sein. Nach dem Unterbruch wurden alle verbleibenden Resultate gleichzeitig übermittelt. Die Frage, ob die ordinale Reihenfolge nach der Unterbrechung tatsächlich das Gleiche misst (was eine Anforderung für die Validität der Methode wäre), wird im Bericht der OAS nicht angesprochen.

Graben entspricht, der sich auch im Referendum manifestierte. Noch wichtiger aber ist, dass unter Berücksichtigung dieser bekannten territorialen Konfliktlinie die Mehrheitsverhältnisse in den später eingetroffenen Resultaten der Präsidentschaftswahlen perfekt erklärt werden, es gibt also keinerlei statistische Hinweise darauf, dass sich während des Unterbruchs, oder in den betreffenden Wahllokalen etwas Besonderes ereignet hat. Dass bolivianische Wahlen in der Regel mangelbehaftet sind, bleibt selbstredend unbestritten, aber der zentrale Indikator für den Betrugsvorwurf der OAS ist damit entkräftet, und die Methode blossgestellt. Über die Gründe, warum sich die OAS von dieser in der Fachliteratur unbekanntem Methode der Wahlforensik überzeugen liess, lässt sich nur spekulieren, über den Hergang ist mir nichts bekannt, auch nicht, ob er intern untersucht wurde.

Manchmal wollen auch autoritär regierte Länder selbst das Vertrauen in ihr Wahlverfahren stärken, etwa Aserbaidschan. Seit der Unabhängigkeit waren alle Wahlen gefälscht und unfrei, alle OSZE-Wahlbeobachtungsmissionen kommen zu einem ernüchternden Bild. 2008 schritt die Wahlkommission zu technischen Hilfsmitteln, und installierte Videokameras in 500 von 5'359 Wahllokalen, um Transparenz im Wahlprozess für die Präsidentschaftswahlen 2008 zu schaffen. Auch das Referendum im Folgejahr wurde in manchen Lokalen direkt ins Internet übertragen; das Land befand über eine Verfassungsänderung um Ilham Aliyev, dem Sohn des früheren Präsidenten Heydar Aliyev, eine dritte Amtszeit zu ermöglichen. Durch die Videoüberwachung sind verbreitete Praktiken, wie etwa „Ballot Box Stuffing“, also das Auffüllen der Wahlurnen mit falschen Wahlzetteln, nicht mehr möglich. Erik Herron hat die Resultate aus Wahllokalen mit und ohne Webcams genauer analysiert.⁴³ In videoüberwachten Wahlstationen unterscheiden sich die Resultate signifikant: die entsprechenden Lokale rapportierten eine um 1.2% geringere Zustimmung für das Referendum, bei einer um 3% tieferen rapportierten Stimmbeteiligung. Ob es sich um einen direkten oder indirekten Effekt handelt, kann nicht eruiert werden. Ein direkter Effekt würde bedeuten, dass in videoüberwachten Wahllokalen Formen des Wahlbetrugs, die zu einem künstlichen Anstieg der Beteiligungszahlen führen, nicht länger möglich waren – womöglich fürchteten die lokalen Wahlkommissionen, dass das Auffüllen des Wahlbehälters mit Stimmzetteln in der Videoübertragung erkannt würde, oder aber sie wollten keine Wahlbeteiligung rapportieren, die von der tatsächlichen – in der Videoübertragung beobachtbaren – Zahl der Wählerinnen und Wähler abweichen würde. Ein

43 E. S. Herron, "The effect of passive observation methods on Azerbaijan's 2008 presidential election and 2009 referendum," *Electoral Studies* 29, 2010, S. 417.

indirekter Effekt läge vor, wenn ein kleiner, aber signifikanter Teil der Wählerinnen und Wähler in kameraüberwachten Wahllokalen ihr Wahlgeheimnis als nicht mehr gewährleistet sähe, und der Abstimmung deshalb fernblieb. Für die Präsidentschaftswahlen stellte Frederik Sjöberg demgegenüber auffällige Abweichungen bei der letzten Ziffer der übermittelten Wahlresultate fest, allerdings nur in den videoüberwachten Wahllokalen (das Referendum hatte er nicht untersucht).⁴⁴ Ein namhafter Teil der von dort übermittelten Resultate ist also nicht gezählt, sondern erfunden. Das passt zur Vermutung, dass das Aufstellen der Kameras die Formen der Manipulation geändert – denn den Wahlhelferinnen und Wahlhelfern unter Videobeobachtung waren die Hände gebunden –, aber nicht zum Verschwinden gebracht hat. Stattdessen hat sich die Wahlmanipulation verlagert, und findet nun während der Auszählung oder Übermittlung der Resultate statt.

F. Schlussbemerkungen

Viele der betrachteten Verfahren sind ausgeklügelt, beruhen etwa auf der Randomisierung der Stimmenden auf Wahlmaschinen. Sie zeigen einerseits, wie sehr die Wahlforensik die Details des Abstimmungsprozesses in die Definition der Methode einbeziehen muss, und andererseits die Grenzen der Übertragbarkeit der Methoden in andere Kontexte.

Eines haben aber alle Methoden gemeinsam: sie belichten nur einen kleinen Ausschnitt aus den vielen Schritten, die für faire und korrekte Wahlen und Abstimmungen notwendig wären, und in aller Regel auch nur eine ganz spezifische Form des Missbrauchs. Weil aber illegitime Eingriffe in jedem der Schritte erfolgen können, bleibt die Möglichkeit mittels statistischen Vorgehens die Qualität einer Wahl oder Abstimmung zu beurteilen, sehr beschränkt und lückenhaft, zumal für die meisten Schritte gar keine Tests bekannt sind. Selbst aufgrund der allgemeinen Natur statistischer Tests gilt: sie schlagen nur bei sehr deutlichen und gehäuft vorkommenden Abweichungen aus, umso deutlicher wenn mehrere Tests parallel durchgeführt werden (und dies korrekt in der Berechnung der Signifikanz berücksichtigt wird). Das heisst, dass der ausbleibende Warnhinweis (ein negatives Resultat) nie als Hinweis darauf verwendet werden kann, dass eine Wahl oder

44 Sjöberg, *Autocratic adaptation: The strategic use of transparency and the persistence of election fraud* (Fn 20).

Abstimmung korrekt durchgeführt wurde; selbst dann nicht, wenn die streitbare Manipulation eigentlich einem Muster entspricht, auf das der Test zugeschnitten wäre.

Tests lassen sich auch standardisieren – mit den erwähnten Vorbehalten. Der Schweizer Kanton St. Gallen etwa verwendet eine Software des Zentrums für Demokratie Aarau (ZDA), um noch während der Erfassung der Teilresultate am Abstimmungssonntag Hinweise auf mögliche Unregelmässigkeiten zu generieren. Hierzu werden automatisch Abweichungen – etwa in der Stimmbeteiligung – zwischen verschiedenen Vorlagen analysiert, oder es wird mittels Vergleichsvorlagen aus der Vergangenheit die Plausibilität der übermittelten Resultate überprüft. Bei solchen Abweichungen handelt es sich allerdings nicht zwingend um Wahlfälschungen. Gemäss der Staatskanzlei des Kantons St. Gallen (mündliche Auskunft) handelt es sich meistens um lapidare, aber folgenschwere administrative Fehler. Insbesondere kann es vorkommen, dass lokale Wahlbehörden die Ja- und Nein-Stimmen verwechseln, aber auch, dass ein Teil der Wahlzettel bei der Auszählung einer Vorlage nicht berücksichtigt wurde, wodurch die Stimmbeteiligung zwischen mehreren parallelen Urnengängen auseinanderklafft.

Es ist allerdings fraglich, ob standardmässige Tests geeignet sind, willentliche Manipulationen von Abstimmungen und Wahlen zu identifizieren. Sobald die Art der Tests bekannt ist,⁴⁵ wird eine professionelle Wahlbehörde – falls sie denn unredliche Absichten haben sollte – darauf reagieren, und die Resultate dahingehend bearbeiten, dass sie nicht weiter auffallen, etwa durch statistische Randomisierung manipulierter Zahlen. Die von Wissenschaftlerinnen und Wissenschaftlern gerne betriebene Wahlforensik leidet unter dem gleichen Problem: publizierte Testverfahren werden auch Wahlbehörden bekannt, zumindest denjenigen mit Fachkenntnissen. Falls sie zudem die Reputation ihrer Arbeit im Auge haben, können sie allfällige Wahl- und Abstimmungsfälschungen vor dem Auge der Wissenschaft tarnen. Damit verliert jedes neu entwickelte Testverfahren mit der Publikation sogleich einen Teil seiner Wirkung.

45 Beispielsweise A. Hicken und W. R. J. Mebane, *A Guide to Election Forensics* (Ann Arbor (MI): University of Michigan), 2015.