



# Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany

Christian Breunig<sup>a</sup>, Achim Goerres<sup>b,\*</sup>

<sup>a</sup> Department of Political Science, University of Toronto, Canada

<sup>b</sup> Faculty of Management, Economics and Social Sciences, University of Cologne, Venloer Str. 153, 50672 Cologne, Germany

## ARTICLE INFO

### Article history:

Received 16 March 2010

Received in revised form 11 March 2011

Accepted 30 March 2011

### Keywords:

Fraud  
Benford's law  
Germany  
Bundestag

## ABSTRACT

This article investigates electoral irregularities in the 1990 to 2005 Bundestag elections of unified Germany. Drawing on the Second Digit Benford Law (2BL) by [Mebane \(2006\)](#), the analysis consists of comparing the observed frequencies of numerals of candidate votes and party votes at the precinct level against the expected frequencies according to Benford's Law. Four central findings stand out. First, there is no evidence for systematic fraud or mismanagement with regard to candidate votes from districts where fraud would be most instrumental. Second, at the state level (*Bundesland*), there are 51 violations in 190 tests of the party list votes. Third, East German states are not more prone to violations than Western ones. This finding refutes the notion that the East's more recent transition to democracy poses problems in electoral management. Fourth, a strong variation in patterns of violation across *Bundesländer* exists: states with dominant party control are more likely to display irregularities. The article concludes by hypothesizing and exploring the notion that partisan composition of nominees involved in the counting may produce a higher likelihood of violation and be a cause of *Länder* variation. This may especially be the case when a party dominates in a *Bundesland* or opponents to the former socialist regime party are involved in the counting.

© 2011 Elsevier Ltd. All rights reserved.

"It's not the voting that's democracy, it's the counting."  
(Tom Stoppard, 1972, *Jumpers*, Act I)

"Electoral manipulators from Dachau [in Bavaria] have to pay more than 116,000 Euro." ([Süddeutsche Zeitung, 2006](#))

Intentional electoral fraud and unintended electoral mismanagement are commonly associated with recently democratized or authoritarian regimes. International institutions such as IDEA in Stockholm or the Office for Democratic Institutions and Human Rights of the OSCE monitor elections in burgeoning democracies in order to ensure that they are fair and credible. However voting irregularities are not limited to those regimes, there is an intensifying

discussion in political science about measuring, detecting, and avoiding electoral irregularities in established democracies. The ballot, voting machine, and counting problems emerging during the 2000 presidential elections in the United States motivated scholars to engage in this field ([Alvarez et al., 2008](#); [Wand et al., 2001](#)). The reason for this heightened interest lies in the fact that democratic qualities are in principle about the adequacies of its processes, especially the vote aggregation procedure.

This article concentrates on electoral forensics of established democracies. Specifically we examine electoral irregularities in unified post-war Germany. Among established democracies, Germany's unification presents a theoretically and empirically interesting case. A formerly socialist, authoritarian regime merged with a functioning electoral democracy, succumbing to the electoral rules of the latter. In 1990 many East Germans were allowed to cast

\* Corresponding author.

E-mail addresses: [c.breunig@utoronto.ca](mailto:c.breunig@utoronto.ca) (C. Breunig), [agoerres@uni-koeln.de](mailto:agoerres@uni-koeln.de) (A. Goerres).

their vote in a free election for the first time. This historical development serves as a natural experiment in which an experienced system of elections expands into a new region which had no prior experience of free elections. If political socialization of the electorate and the experience of organizing elections matter, distinct patterns in the procedure of aggregating voting should have emerged for West and East Germany after 1990.

Unified Germany is not known for electoral fraud or mismanagement. Indeed, in the first outside observation of a German Bundestag election, the OSCE in 2009 (OSCE, 2009: 24) reported that because of “high level[s] of trust in the election administration by most parties and the public, there is generally little interest in observing the process.”<sup>1</sup> Consequently, international institutions refrain from conducting systematic monitoring during German elections. However, individual cases of electoral fraud did occur in the recent past, such as the case quoted above of the local election in Dachau Bavaria in 2002. Two politicians of the Christian-Social Union (CSU) manipulated hundreds of ballots and acted as proxies for dozens of friends casting the ballots in their own interests. Moreover, since unification (see Fig. 1), private citizens have substantively increased appeals to the German Bundestag against Bundestag elections. The growing number of petitions are an indicator of rising suspicion for electoral irregularities by the electorate (see Ziblatt, 2009). Before 1990, the number of appeals never exceeded 50 per parliamentary session; their numbers grew after 1990 with a massive outlier of 1453 in the 1994–8 session.

In this article, we employ recently developed statistical techniques in order to investigate electoral irregularities. Specifically, we examine whether patterns of electoral returns match the theoretical expectations based on the Benford distribution. Statistically significant deviations from the distribution point us toward electoral irregularities that can be due to either mismanagement and/or fraud. The technique does not allow us to differentiate between the two types of irregularities. Like a thermometer, it reveals that something is amiss, akin to a fever, without giving the underlying cause. However, if irregularities are detected, there is an arguably strong incentive to conduct further enquiries into potential problems. We apply this method to both votes that are cast in federal elections: the first vote for a local candidate who gets elected by simple plurality (*Erststimme*) and the second vote for a closed party list that is counted under a system of proportional representation with a nationwide threshold (*Zweitstimme*). The data employed is the lowest aggregated vote count data available, namely from precincts (*Wahlbezirke*) with each precinct having a maximum of 2500 eligible voters.

Section 2 reviews the existing studies of electoral irregularities, including fraud and mismanagement. Section 3 presents the data and the techniques of the empirical

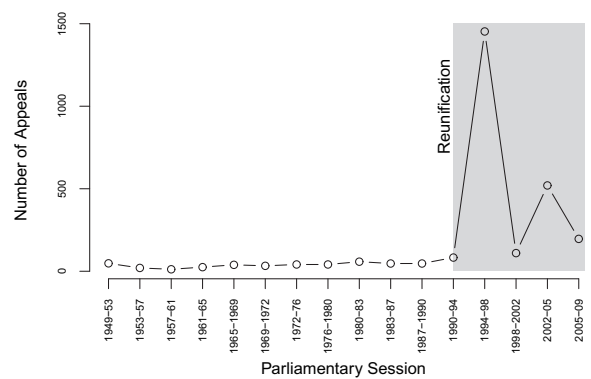


Fig. 1. Number of appeals by private citizens against Bundestag elections, 1949–2005. Sources: Deutscher Bundestag (2000; Deutscher Bundestag, 2009a), Feldkamp and Ströbel (2005).

analysis. Section 4 describes the Benford test results and outlines some empirical exploration of the patterns found. The final section evaluates the results. Most importantly, there is no evidence for irregularities with the candidate votes. This is reassuring since the manipulation of candidate votes in a district would enable the possibility of successful fraud for one seat in the *Bundestag*. We find evidence for irregularities in the party list votes. We speculate that the pattern of irregularities might suggest problems with the organization of the counting process and the composition of the group of appointed electoral officers.

## 1. Current research on electoral irregularities

There are three types of analyses to detect electoral irregularities. First, process-based study of particular elections rely, for example, on evidence from witness accounts of vote-rigging (see for a list of references Lehoucq, 2003). Second, aggregate results can be analyzed via regressions (Wand et al., 2001). Here, researchers construct regressions with the use of suitable covariates in order to predict the district level fortunes of parties or candidates. Statistical outliers are identified as unusual cases and subjected to case studies. Third and most important for this article, there are studies of aggregate results that check deviation of digits of electoral results from distributional laws that hold for certain digits. Several studies employ this strategy including analyses of 19<sup>th</sup> century US presidential and gubernatorial elections (Buttorff, 2008), of US presidential elections in 2000 and 2004 (Mebane, 2006, 2007), of Mexican elections (Mebane, 2006), of Swedish versus Nigerian elections (Beber and Scacco, 2008), Taiwanese elections (Chiang and Ordeshook, 2009), Russia and Ukraine (Myagkov et al., 2009). The 2009 Iranian election generated a lively scholarly discussion after an initial analysis of digit distributions of vote returns (Roukema, 2009; see for example Gelman, 2009; Beber and Scacco, 2009; Freiberger, 2009; Mebane, 2009).

Mebane (2006) proposes the most elaborate model of how to study the distributions of digits of vote returns. He suggests the “Second Digit Benford’s Law (2BL)” and gives explicit mechanisms for the expectation that vote count

<sup>1</sup> The mission consisted of a team of 15 election experts from 13 OSCE countries. They visited Germany twice, once before the election for three days in July, and then for 19 days around election time. They went to all 16 states. They did not conduct observations of counting except for six polling stations.

data follow Benford's Law for the second digit. Benford's Law describes a numerical regularity in real-world numbers expressed in the decimal system. According to Benford's Law<sup>2</sup> (Benford, 1938), various digits in lists of numbers do not occur with the same frequencies. For example, the formulas for the first digit is  $P(d_1) = \log_{10}(1 + (1/d_1))$  with  $d_1$  being one of the numerals 1,...,9. More generally, the joint distribution of the first and all later digits adhere to the following probabilities (Hill, 1995):  $P(D_1 = d_1, \dots, D_k = d_k) = \log_{10}[1 + (\sum_{i=1}^k d_i 10^{k-i})^{-1}]$  with  $D_1, D_2$  signifying the first, second etc. significant digit and  $d_i$  the numerals 0, 1,...,9 ( $j = 2, \dots, k$ ).

Table 1 shows the expected frequencies according to Benford's Law for the first to the fourth digits. In essence, the expected frequency of numerals is skewed for the first digit and approaches a uniform distribution for higher-order digits. It has been shown that a Benford distribution of frequencies is expected to occur (a) if numbers stem from several natural processes that create individual distributions of numbers, (b) if the processes bringing together the numbers are not selected in a systematic way, and if (c) the numbers are representative of the distribution they come from (such as a random sample of numbers from each distribution) (Hill, 1995). For example, if someone copies down all numbers from the front page of a daily newspaper and does that for a year, the resulting distribution of numerals in that distribution approximates the Benford distribution (Hill, 1998). In this case, the numbers come from several distributions (page numbers, sport results, daily temperature, stock market indices etc.), and the distributions are not systematically selected; they just occur naturally together on the first page of the newspaper.

Mebane (2006) argues that the frequencies of the numerals of election counts at precinct level approximate a Benford distribution for the second digit. They must do so because the aggregate distribution of numerals does not result from one random process. Instead, it stems from a mixture of several intersecting processes (such as multiple decisions whether to vote or not, decisions among a given set of menus that vary across districts, voting technology etc.) (see also Mebane, 2008). Other researchers have used the first digit of election results (Roukema, 2009); however Mebane rightly argues that the first digit cannot vary enough e.g., if a candidate benefits from similar popularity in precincts of similar magnitude. The second digit provides us with the opportunity to employ Benford's Law for studying voting irregularities. In fact, Mebane's empirical tests with various kinds of empirical data from the USA, Mexico, Iran and with artificial data simulating fraudulent elections corroborated the 2BL.

## 2. Testing the Second Digit Benford Law on German Bundestag elections

The advantages of aggregate-level analyses drawing on distributional laws are threefold. Compared to regression-based analyses, there is no need to have covariates

**Table 1**

Probabilities of first and higher-order digits as predicted by Benford's Law.

Numeral	Probabilities of numerals 0–9			
	Digit 1	Digit 2	Digit 3	Digit 4
0		0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.10882	0.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.0994	0.09994
7	0.05799	0.09035	0.09902	0.0999
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982

(Mebane, 2006: 1). This reduces demands in terms of data and theory (e.g. for turnout and voting preferences). In addition, digit tests can be performed rather quickly. For example, Beber and Scacco (2009) using another distributional test suggested that Iranians official results from the 2009 presidential elections had been manipulated only eight days after the election had been held. Furthermore, cheating “successfully” requires cheaters to have access to all numbers at the same time (a very high hurdle) because it is, of course, possible to generate faked results following a Benford distribution when knowledge about these systemic regularities of digits is available.

However, we should be clear that second digits tests do not discriminate between error and fraud and whether a party of candidate has won or lost because of the uncovered irregularities. Thus, the technique ends in an inductive interpretation of the patterns of violations and can inform other follow-up techniques, such as the selection of case studies of individual districts. These tests resemble a thermometer indicating fever: the origins must be sought by other diagnostics.

Based both on our literature review and an interview with a representative of the German federal returning officer, we are not aware of any analyses of irregularities of the electoral process in post-War Germany. This omission is striking because it reveals either a lack of interest in the community of electoral scholars and/or a lack of concern by practitioners. One explanation may be that there is a very sound legal framework that, by federal decree, meticulously prescribes every detail of the vote-counting process (see below). The counting process is very complex and requires the involvement of four levels of administration (federal, *Land*, electoral district and electoral precinct). This transparency seems to represent a seemingly sound system of checks and balances.

Moreover, elections in Germany are rarely questioned before the Constitutional Court.<sup>3</sup> Basically, any citizen can file a petition up to two months after an election and question the validity of the last election (see Fig. 1). No complaint has ever been accepted by the *Bundestag*, meaning that the *Bundestag* has never questioned the

<sup>2</sup> Actually, the first one to publish on this regularity was Newcomb (1881).

<sup>3</sup> On 3 March 2009, the Federal Constitutional Court took a decision, prohibiting the usage of computers for general elections. The Court also ruled on the so-called negative weight votes and as a consequence of this decision the *Bundestag* is forced to change electoral rules by 2012.

validity of an election. Based on these observations, one might conclude that there is no indication that voting irregularities occurred in German *Bundestag* elections. Moreover, individual incentives for purposive manipulation of vote counts at the precinct level might not be that great because electoral outcome for the party list vote (*Zweitstimme*) are based on nationwide proportional representation. Individuals with only local access to the results could only manipulate precinct level returns that, given the mixed-member system, produce at best marginal changes in the overall outcome.

Finally, the complexity of the counting process calls into question whether there is room for human error and manipulation. Especially given the different histories of East and West, we would expect more violations from the East than in the West, particularly in earlier elections. Irregularities might be due to the high organizational requirements of the counting process. These requirements might be more difficult to manage in a region lacking recent experience with free and fair election. In sum, a thorough analysis of potential irregularities at German *Bundestag* elections seems warranted and necessary.

### 3. Data and methods

Our analytical strategy proceeds in three steps. First, we obtain the vote counts for each candidate and all major parties at precinct level. We then calculate the frequencies of all numerals for the second digit on an appropriate subset of these data. The third step checks whether the distribution of all numerals follows the pattern that Benford's Law predicts for the second digit. A simple Pearson's  $\chi^2$  statistic adjusted for multiple testing provides us with the probability of the mismatch between the observed and expected frequencies. Each of these steps warrants a detailed explanation.

The data for our analysis relies on precinct-level results (candidate vote and party list vote) for each German general election between 1990 and 2005. We simply count the occurrence of each numeral at the second digit separately (i) for all candidates from all precincts in a district, (ii) for all parties from all precincts in a district and (iii) for each party individually from all precincts for a *Bundesland*. With the party list votes, we take a conservative approach and restrict the analysis to the major parties the Christian-democratic CDU/CSU, the Social-democratic SPD and the Socialist PDS/Left (in the East). We do this for two reasons. First, the Second Digit Benford Law requires returns to have a second digit in most of the occurrences. This precludes inclusion of small parties, such as the right-wing NPD, but also the Greens and the liberal FDP which were minuscule parties in many precincts in the East.

Second, strategic voting considerations, rather than sincere voting, lead voters vote for a party or candidate other than their first preference in order to influence the outcome (Gschwend, 2007). In Germany, voters can strategically vote by splitting their ballot between a candidate vote for party A and a party vote for party B. In the overwhelming number of cases, districts are won by the CDU/CSU or SPD candidates. Knowing this, voters with a first preference for one of the two parties can cast their party list

vote for a smaller party in order to express their preference for a certain coalition. For example, a CDU partisan votes for the FDP with the party list vote and for the CDU candidate with the candidate vote in order to support the preferred right-of-center coalition between CDU/CSU and FDP. According to the logic of Benford's Law, such considerations make the processes, whose outcomes are the distributions of numerals, no longer independent from each other. The process that leads to the voting decisions is no longer independent of the process of the generation of voting options. We address this problem by focusing on each vote choice and large parties individually. Based on a reviewer's suggestion, we also combined the votes for each *Lager* (i.e. CDU/CSU/FDP and SPD/Grüne/B90) in an additional analysis in order to assess the strategic motivation of voters. The *Lager* results mirror those of the individual party tests and are obtainable from the authors.

We employ the following methodology in order to assess irregularities. We compare the frequency distribution of the occurrence of each numeral at the second digit against the expected Benford distribution using a Pearson's  $\chi^2$ -test with nine degrees of freedom. Because multiple significance tests are conducted simultaneously, we need to adjust for the expected proportion of falsely rejected hypotheses (see Mebane (2006)). This error rate is called the False Discovery Rate (FDR) (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2005).

While Benjamini and colleagues (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2005) provide a formal definition of the FDR, the procedure is easily summarized. When assuming independence across tests and testing  $m$  hypotheses  $H_1, H_2, \dots, H_m$  with the corresponding  $p$ -values  $P_1, P_2, \dots, P_m$ , we can order the  $p$ -values from smallest to largest:  $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$ . The following multiple testing procedure controls the FDR: for a chosen test level  $\alpha$ , let  $k$  be the largest  $i$  for which  $P_{(i)} \leq (i/m) \cdot \alpha$  and then reject all  $H_{(i)}$  with  $i = 1, 2, \dots, k$ . In short, using the FDR procedure avoids the multiple comparison problem by readjusting the expected proportion of incorrectly rejected null hypotheses.

In our analysis we chose the critical  $\chi^2$  value of the .01 level of confidence. The critical value for a  $\chi^2$  distribution with 9 degrees of freedom is approximately 21.7 for the .01 level of significance. This critical value increases when we take the number of individual tests results and the resulting FDR into account. For example (see Table 2), the FDR-adjusted  $\chi^2$  value is 37.87 for the test with CDU/CSU votes in the 1990 election. In case the FDR-adjusted  $\chi^2$  test statistic exceeds the critical values, we can reject the hypothesis that the second digits are distributed according to the 2BL.

Our approach can be illustrated with a simple example. Fig. 2 displays the actual and expected percentages for second digit of the precinct level vote counts for the CSU in Bavaria in the 1990 federal election. In order to obtain this graph, we first collected all votes for the CSU in each precinct. We then extracted the second digit from each vote count and tabulated for each of the percentage of zeros, ones, etc. among the 13,377 second digits. Finally, we computed the difference in expected vs. actual distribution of digits via an adjusted Pearson's  $\chi^2$  test. For the second digit, the figure illustrates that there are fewer zeros, ones,



**Table 2**  
Adjusted Pearson's  $\chi^2$  results for the comparison of expected second digit numeral counts of party list vote with the actual results per party CDU/CSU, SPD, PDS/Left per Bundesland in 1990–2005.

Bundesland	Election/party 1990						1994						1998						2002						2005					
	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left	CDU/CSU	SPD	PDS/Left
West	88.85	32.93		47.92	46.12		43.90	54.46		37.92	42.55		43.90	54.46		37.92	42.55		36.60	33.52		36.60	33.52		36.60	33.52		36.60	33.52	
Baden-Wuerttemberg	42.56	48.09		44.15	50.40		43.22	32.88		76.20	92.46		43.22	32.88		76.20	92.46		50.10	135.51		50.10	135.51		50.10	135.51		50.10	135.51	
Bavaria	6.42	14.65		21.02	13.10		8.17	23.98		8.03	24.19		8.17	23.98		8.03	24.19		8.96	27.02		8.96	27.02		8.96	27.02		8.96	27.02	
Bremen	13.41	9.98		15.00	12.62		11.72	26.43		17.39	12.56		11.72	26.43		17.39	12.56		5.46	9.14		5.46	9.14		5.46	9.14		5.46	9.14	
Hamburg	12.37	38.86		18.05	23.23		17.73	31.24		19.11	38.40		17.73	31.24		19.11	38.40		24.92	26.99		24.92	26.99		24.92	26.99		24.92	26.99	
Hessen	23.42	40.64		39.58	17.92		14.45	26.56		20.43	28.10		14.45	26.56		20.43	28.10		26.03	23.54		26.03	23.54		26.03	23.54		26.03	23.54	
Lower Saxony	47.67	86.65		39.99	62.40		14.88	139.19		25.89	86.22		14.88	139.19		25.89	86.22		40.28	65.84		40.28	65.84		40.28	65.84		40.28	65.84	
North-Rhine-Westphalia	12.73	7.64		16.94	5.70		16.06	10.75		11.96	7.06		16.06	10.75		11.96	7.06		15.43	22.34		15.43	22.34		15.43	22.34		15.43	22.34	
Rhineland-Palatinate	22.80	7.43		10.92	20.53		12.07	23.88		27.37	37.60		12.07	23.88		27.37	37.60		10.45	8.43		10.45	8.43		10.45	8.43		10.45	8.43	
Saarland	17.39	9.18		21.71	20.04		18.49	10.94		9.06	18.53		18.49	10.94		9.06	18.53		12.10	18.21		12.10	18.21		12.10	18.21		12.10	18.21	
Schleswig-Holstein	37.87	18.72	37.57	19.51	32.78	32.32	25.76	37.27	17.57	11.96	32.53	58.38	25.76	37.27	17.57	11.96	32.53	58.38	9.28	40.43	29.24	9.28	40.43	29.24	9.28	40.43	29.24	9.28	40.43	29.24
Berlin	11.95	23.33	27.08	7.82	14.87	9.69	11.07	11.81	15.18	6.84	11.14	34.37	11.07	11.81	15.18	6.84	11.14	34.37	25.23	4.26	9.19	25.23	4.26	9.19	25.23	4.26	9.19	25.23	4.26	9.19
Brandenburg	15.93	12.76	17.79	3.23	7.74	8.44	8.59	6.99	4.23	17.72	5.75	29.45	8.59	6.99	4.23	17.72	5.75	29.45	7.12	13.19	31.75	7.12	13.19	31.75	7.12	13.19	31.75	7.12	13.19	31.75
Mecklenburg-Western Pomerania	27.78	14.31	10.38	9.45	9.56	19.63	17.32	9.09	19.35	15.27	19.96	56.78	17.32	9.09	19.35	15.27	19.96	56.78	22.09	15.06	16.78	22.09	15.06	16.78	22.09	15.06	16.78	22.09	15.06	16.78
Saxonia	7.65	11.72	7.82	5.03	14.55	19.49	4.37	15.13	11.29	12.03	11.43	36.31	4.37	15.13	11.29	12.03	11.43	36.31	8.93	7.38	10.30	8.93	7.38	10.30	8.93	7.38	10.30	8.93	7.38	10.30
Saxonia-Anhalt	13.19	20.01	16.45	10.95	13.59	9.04	11.22	7.60	14.01	7.01	31.96	40.51	11.22	7.60	14.01	7.01	31.96	40.51	22.60	8.58	12.23	22.60	8.58	12.23	22.60	8.58	12.23	22.60	8.58	12.23
Thuringia	37.87	32.93	27.08	39.58	32.78	32.32	43.90	31.24	–	37.92	28.10	29.45	43.90	31.24	–	37.92	28.10	29.45	36.60	33.52	29.24	36.60	33.52	29.24	36.60	33.52	29.24	36.60	33.52	29.24
Critical FDR-adjusted $\chi^2$ at .01 level																														

and two (and more sixes, sevens, eights, and nines) among the vote counts than a Benford distribution would expect. This disparity between the actual and the expected distribution of digits is reflected in a  $\chi^2$  of 42.6 which is clearly above the critical value of 21.7 for the .01 level of significance. As such, the distribution of second digits among the precinct level results for the CSU suggests that counting irregularities may have occurred.

#### 4. Empirical analysis

##### 4.1. Description of the counting process

There are three legal documents that prescribe the voting procedure at German federal elections. The German Constitutional Law (last amended 19 March 2009) describes the general principles of voting (especially Art. 20, 28, 38, 29, 41). Most importantly, it prescribes the general nature of elections (general, direct, free equal, secret), eligibility criteria (German citizens aged 18 and older with some exceptions) and that the German Bundestag itself handles complaints against electoral procedures, decisions that can subsequently be brought before the Constitutional Court. The law of elections (*Bundeswahlgesetz*, last amended 17 May 2008) lays out the legal requirements of General Elections in detail. Finally, a decree (*Bundeswahlordnung*, last amended 3 December 2008)<sup>4</sup> meticulously sets out every detail of the whole process, including, for instance, very specific regulations about how the ballots should be counted and packed together.

As a reflection of German federalism, the voting process is organized in four layers: (1) federal: the federal returning officer (*Bundeswahlleiter*) is a bureaucrat and usually the head of the German Office for National Statistics. This person and a replacing officer are nominated for an indefinite period. The federal returning officer nominates members of a supervising committee and their replacements (*Bundeswahlausschuss*); (2) state (*Land*): the state returning officer (*Landeswahlleiter*) and his/her replacement are nominated for an indefinite period. The state returning officer nominates committee members and their replacements; (3) electoral district: the district returning officer (*Kreiswahlleiter*) and his/her replacement are nominated for one electoral cycle only. The district returning officer selects members for a committee and their replacements; (4) precinct: precinct returning officers, their replacement and their assistants (*Beisitzer*) are nominated for one election only. For postal votes and mobile voting stations (such as for hospitals), additional regulations apply.

All members of committees and their replacements have to be eligible to vote and should be resident in the region for which they serve. Political parties are able to recommend officials based on the order of electoral success in the respective region at the last election. Citizens cannot simply refuse to be called upon for duty. They are excused if they are

<sup>4</sup> The *Bundeswahlordnung* is the source of information for the following paragraphs.

members of federal or state governments or of the European, federal or state parliaments, older than 64 or ill, have care responsibilities or can give “plausible” professional reasons. If parties do not make recommendations or those recommended are not available, no further official regulations exist.

No-shows at election day at precinct level can be replaced by voting citizens. Anecdotally, we know that many local communities ask for volunteers and use random-selection procedures to get a sufficient number of helping hands. The *Bundesland* of Berlin ordered all teachers, who are civil servants of the *Land*, to act in these committees for the 2009 European election and a regional referendum (Schulz, 2009). Once called upon, individuals seem to be likely to be asked again and to become promoted from assistant to precinct returning officer or replacement. In total, about 630,000 people are working on election day in these various functions (Deutscher Bundestag, 2009b). So far, no empirical study exists of these procedures and the likely social selection bias that we could expect, namely that those who are known to be politically active, longer residents and those who have done it once are likely to be called for duty.

All told, the vote collection and counting procedure are considerably complex and involve a number of lay and professional personnel. This process, on the one hand, allows for human error and, on the other hand, makes purposive manipulation difficult because there are many checks and balances. Further details on the procedures can be found in the appendix.

#### 4.2. Candidate vote counts

The first step of the test procedure involved conducting approximately 1500 tests in about 300 districts across five elections to test for deviations from the 2BL for the candidate vote counts. Let us illustrate the procedure for one concrete example, the Bavarian electoral district Hof (ID # 240) in the 2005 elections. We first identify all regular polling stations, i.e. not mobile or virtual postal ones, in the district and then select the electoral returns of candidates that received at least 10 votes in one precinct. For the Hof example, this procedure returns 1372 useable precinct-level counts that stem from the district's 331 polling stations and includes vote counts from candidates of the SPD, CSU, GRÜNE, FDP, Linke, and NPD (in this case: all candidates). From the 1372 counts we extract the second digit and count the number of 1's, 2's, ... and 0's. The observed frequencies of the numerals are compared to their expected frequencies according to 2BL. This test delivers a  $\chi^2$  test statistic of 46.1 that is clearly above the .01 threshold of 21.7. An adjustment for multiple testing is unnecessary since we only conduct one test on that district and these candidates.

In total, there are only four violations that are statistically significant at the .01 level, a number that we would expect due to chance alone: Hamburg-Nord (Hamburg) and Pinneberg (Schleswig-Holstein) in 1994, Deggendorf (Bavaria) in 2002 and our example, Hof (Bavaria) in 2005. In these electoral districts, the counts of numerals at the second digit from all candidates at the precinct level

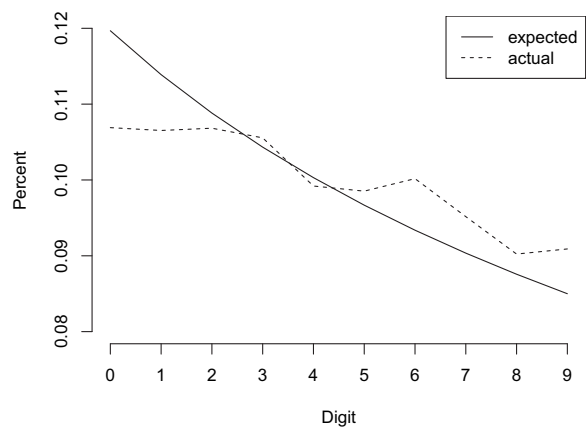


Fig. 2. Illustrative example of second digit count results versus expectations, party list votes, CSU, Bavaria, 1990.

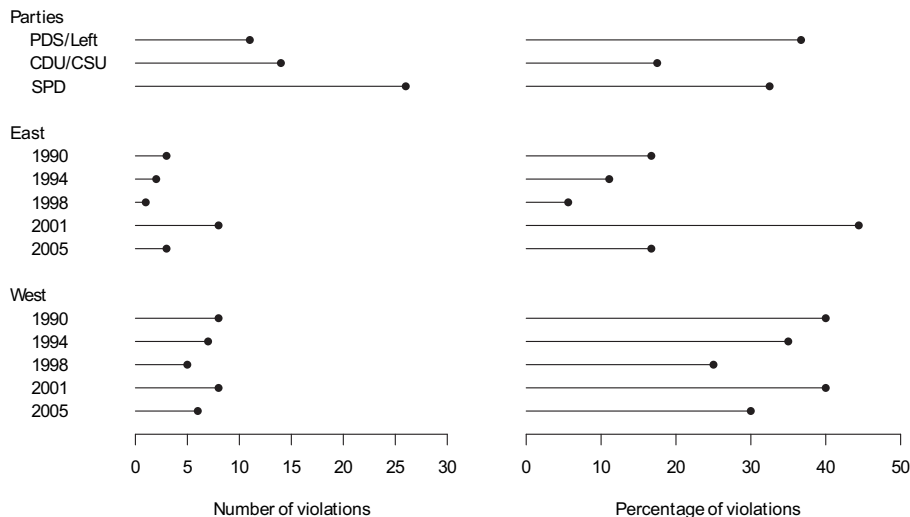
deviate so substantively from the expectations according to Benford's Law that it is highly improbable that chance alone can explain these four deviations. However, given the large number of tests, the overwhelming number of non-violations – implying a regular electoral process – is very impressive.

At the district level, individual candidates only need a plurality of votes to win a seat in the *Bundestag*. Thus, fraud is actually tempting and instrumentally sensible at the district level. If a perpetrator managed to rig the results, his or her candidate would win that district. The absence of a sizable number of 2BL violations means that there is no substantial evidence for easily manageable electoral fraud in unified Germany.

The four violations could be followed up by case studies in order to find district evidence of irregularities. Some of the following procedures might deliver additional insights into the four cases and might enable researchers to assess whether these violations are due to error or fraud. A straightforward first step would be checking the archived summary documents of the counting in the federal archive in Koblenz. If returning officers had been suspicious about the way in which the counts had been produced, they could have left a signed note about their suspicion in these documents as proscribed in the decree. In addition one might interview the local returning officers, which is of course easier for more recent elections, in order to elucidate where they noticed any irregularities. Finally, one could compare the precinct results of the year with a violation with a precedent and later set of precinct results to see whether certain party results deviated from long-term results in these precincts using a regression type analysis as outlined above.

#### 4.3. Party list counts

As a follow-up to the candidate level results, we aggregate the returns for all parties with the party list-vote per voting district and conducted another 1500 tests on these aggregates. Once again, there is very little evidence of electoral discrepancies; we identify irregularities in only nine cases: Essen III (North-Rhine-Westphalia) in 1990, Freiburg (Baden-



**Fig. 3.** Number of 2BL violations per party and in East/West per election year, absolute numbers and relative proportions in %. Note: a total of 190 tests were carried out per party on the level of the *Bundesland*: SPD, CDU/CSU ( $2 \times 16 \text{ Länder} \times 5 \text{ elections}$ ) + PDS/Left ( $6 \text{ Länder} \times 5 \text{ elections}$ ). Percentage points are always relative to the number of tests carried out for that category. For example, 26 of the SPD tests out of 80 carried out for that party had significant deviations (=32.5%).

Wuerttemberg), Hof, Kulmbach, (Bavaria), Frankfurt/Oder – Eisenhüttenstadt – Beeskow (Brandenburg) in 1998, Rostock, Stralsund – Nordvorpommern – Rügen (Mecklenburg-Western Pomerania) in 2002 and Amberg (Bavaria) in 2005. Note, however, that the voting district of Hof in Bavaria appears both in the group of suspect districts in 1998 here and for the candidate votes in 2005. These results might justify a detailed case analysis of counting process in Hof. Overall, given the high number of tests, chance alone would have brought about a few significant violations.

We now consider 2BL tests for vote counts of individual parties – the CDU/CSU, SPD and the PDS/Left (only East) – aggregated from all precincts for each *Bundesland* (see details in Table 2). At this level of aggregation, we identify more violations than the district level tests indicated: there are 51 potential violations out of 190 tests. Fig. 3 summarizes the number and relative proportion of tests showing significant violations. These violations can be explored along three dimensions: parties, region and time.

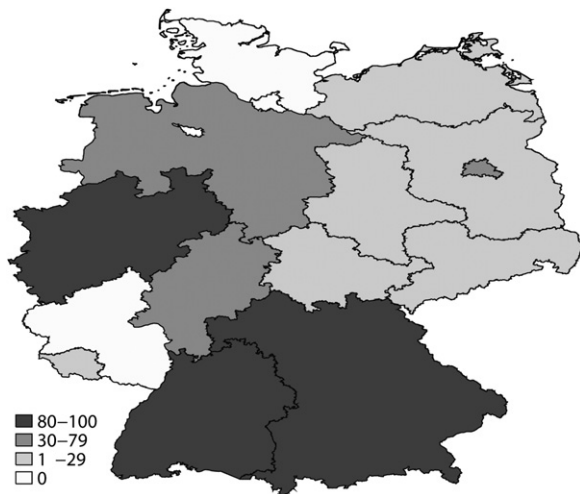
First, among the parties, the results for the Christian-Democratic Party show fewer violations (17.5%) than for the SPD (32.5%) or PDS/Left (36.7%). In addition, PDS/Left violations (11) are more common in East Germany than SPD (5) and CDU (1). If these violations indicate irregularities that work against or in favor of a party, the SPD and PDS/Left seem to be affected more than the CDU/CSU. However, using our methodology and the accompanying results, it is not possible to ascertain whether and which parties suffered or gained from the uncovered irregularities.

Second, it might be expected that election returns in the East show more violations of the 2BL tests than in the West, especially in earlier elections. A higher occurrence might indicate the absence of experience in the grass-root level organization of elections. In the Eastern states alone, more than an estimated 100,000 people are required to be involved in the counting process on election day. Given the complexity of the rules stated in the decree guiding the volunteers, it would come as little surprise to find more

violations in those areas where fewer people with experience have been involved in the counting process. However, Fig. 3 refutes this expectation by showing that with the exception of 2002, the relative proportion of significant deviations from 2BL was lower in the East than in the West. Thus, there is little evidence for the assertion that, in general, the Eastern states are more prone to violations than the Western. The high number of deviations in the 2002 election for the East is due to violations of the PDS/Left tests exclusively. This finding is remarkable and raises some concerns because the PDS/Left was expected to be very close to missing the nationwide threshold of 5% in 2002. The substantial instrumental incentive to manipulate the second vote in the East in 2002 combined with the high number of violations of the 2BL for the PDS/Left in that election warrant thorough inquiry.

Third, it is useful to explore the test results by *Bundesland* (see Fig. 4). The map<sup>5</sup> highlights four different levels of violations. For each *Bundesland*, we calculate the proportions of tests revealing violations. Thus, for Berlin, for example, there is a total of 15 tests (three parties SPD, CDU and PDS/Left times 5 elections), 9 of which (60%) showed violations. Four *Bundesländer* display absolutely no violations: the city-states Bremen and Hamburg, Rhineland-Palatinate and Schleswig-Holstein. Six *Bundesländer* have a proportion of 6.7–13.3% of their tests showing violations of 2BL. These are all East German *Länder* plus Saarland. Three *Länder* appear to violate between 30 and 60% of the tests, namely Berlin, Lower Saxony and Hessen. North-Rhine-Westphalia (80%), Bavaria (90%) and Baden-Wuerttemberg (100%) show the highest relative number of violations. The high number of violations for the three *Bundesländer*, North-Rhine-Westphalia, Bavaria and Baden-

<sup>5</sup> The map was drawn with the *smap* STATA command written by Maurizio Pisati. The map shapefile is taken from the GADM database of Global Administrative Areas ([www.gadm.org](http://www.gadm.org)).



**Fig. 4.** The geographical distribution of the violations of the Second Digit Benford Law, 1990–2005. Note: Shades denote the relative proportion of all tests carried out per *Bundesland* that showed violations of the 2BL at 0.01 level. For example, in Baden-Wuerttemberg, 10 tests were carried out (two parties CSU and SPD times 5 elections = 10 tests) and all of them (100%) showed violations.

Wuerttemberg, is truly striking. The implication is that almost all party list results for the two biggest parties, SPD/CDU in NRW, CSU/SPD in Bavaria and CDU/SPD in Baden-Wuerttemberg violate the expectations in the second digit according to Benford's Law.

#### 4.4. Exploring some patterns of violation

What might explain the described pattern of violations? A distributional analysis based on Benford's Law cannot help us to answer this question. Here, we provide some initial speculations about possible causes and delineate some exploratory analyses. We explore four avenues of inquiry focusing on demographic size, regional party dominance, social homogeneity, and recruitment problems in Eastern Germany.

First, the more people live and vote in the state, the higher the number of violations per *Bundesland*. This occurrence is at least partially a derivate of the statistical test. The association between demographic size and the number of violations partly reflects the high number of numerals that are summed up in a state like North-Rhine-Westphalia with about 18 million inhabitants compared to the smaller amount of numerals in a state like Bremen with less than 1 million inhabitants (Statistisches Bundesamt, 2009: chap. 2). Differences in the size of votes then influence the Pearson  $\chi^2$  test due to its sensitivity to sample size. However, this sensitivity does not explain why Berlin has 60% of tests showing violations whereas Rhineland-Palatinate has 0% even though both have about 3.5 million inhabitants and a similar number of voting precincts. Also, overtime variations would be hard to explain. For example, the  $\chi^2$  values for the SPD in Saarland vary between 7.43 and 37.60 across elections, even though the number of precincts remains nearly identical. In addition to the statistically derived influence of population size, population size may

also be relevant for the counting process as such. The more eligible voters there are in a state, the more precincts are in a state and the greater the organizational burden and thus the proneness to human error.

A second potential source of the geographic clustering of irregularities may have to do with the regional dominance of a party in a state and its associated capacity to recruit counting volunteers. Specifically, the three states with frequent violations are also known for its dominant parties: the CSU in Bavaria, the CDU in Baden-Wuerttemberg and the SPD in North-Rhine-Westphalia. In the 1990 election for instance, the CSU obtained 51.9% of all party list votes in Bavaria and its sister party, the CDU in Baden-Wuerttemberg obtained 49.5% of all party list votes whereas the CDU received as little as 31.1% in Bremen. The SPD scored 38.4% in North-Rhine-Westphalia, which was not its best result. However, within North-Rhine-Westphalia, the SPD received as much as 60.0% in the district of Duisburg II. Partisan dominance at any level, ranging from *Bundesland* to precinct, may lead to a selection bias for volunteering. If voting is dominated by a specific party in a region, then volunteers also are likely to reflect this partisan bias. This dominance, in turn, may lead to distortions in vote counting. Returning officers who share a partisan affiliation might be less vigilant and therefore more likely to commit more mistakes while counting.

A third cause could be varying levels of social homogeneity within the populace across *Bundesländer*. At the micro-level, appointees working together could be more homogenous in some contexts than in others. More social homogeneity might breed less control in counting as group members could be less distrustful of each other or be less willing to control someone like themselves. This proposition can be grounded on the sociological literature on homophily, the human tendency to be attracted to people like oneself and to be affected in behavior and attitudes by this tendency (McPherson et al., 2001).

A fourth possibility is that the recruitment process of the returning officers in the Eastern states is problematic. We showed that the PDS/Left results violate 2BL more often than those of the SPD and CDU in the East. This variance might be attributed to some construction of both intentional fraud and unintentional errors by a selected group of appointees. In the East, the PDS/Left is the successor party of the SED, the former state party of the GDR. In the earlier elections, those partisans might be less likely to volunteer and be selected as election officers as local officials want to break with the past. These dynamics might shape a group of returning officers in the East that is marked anti-PDS/Left. Anecdotal evidence suggests that many communities utilize civil servants in order to conduct the counting in early elections, such as in the 2009 Bundestag elections in Berlin. From March 2009 until 2014, an administrative decree states that civil servants of the *Bundesland* Berlin get one day off if they have helped in the election (Landesregierung Berlin, 2009).

Again, elections officers might be more likely to manipulate against the PDS/Left or might be less vigilant and produce more human errors as a group where these types of selection bias prevail. The 2002 violations for the PDS/Left seem to be particularly intriguing. In 1998, the PDS just



made it into the Bundestag with 5.1% of all votes. In 2002, its vote share did not cross the 5% threshold. With 4.0% of the vote, the PDS sent only two MPs with a direct mandate from their respective districts in the *Bundestag*. The high number of violations at state level (6 out of 6) might be attributable to an intentional attempt to change the result of the PDS/Left for it not to cross the electoral threshold. In short, there could have been a stronger instrumental incentive to cheat against the despised former regime party in 2002. We believe that these observations especially are worth exploring more carefully in future research.

Using a subset of our voting data, we are able to explore the first three propositions in an analysis of the 2005 party list results for *Bundesländer* with a substantial number of violations. By the 2005 election, Germany was “most unified”, i.e. observations across the East and West are fairly homogenous.<sup>6</sup> Our analysis has to remain exploratory because we only examine 16 observations, and measurements of the independent variables are rudimentary. Most importantly, due to the multiple-test design of the distributional analysis, the dependent variable – the  $\chi^2$  values – does not have a fixed meaning. In other words, a test value of 25 in one case can be a violation at the .01 level, in another case it may not be. In short, this exercise serves as generating some plausible suggestions and pointing toward empirical tools of testing them.

We regress the  $\chi^2$  test results on three variables. First, we use population size in millions (ranging from 0.66 to 18.06) (*Statistisches Bundesamt, 2010a*). Population size is used in order to control for the test-inherent sensitivity to the number of precinct observations, which correlates highly with population size. The theoretically more interesting hypothesis is that in larger populations the counting process occurs across a larger number of precincts and thus is more likely to produce a higher number of human errors. We thus expect a positive sign of the coefficient of population size.

Second, we attempt to capture social homogeneity using a measure based on in and out migration relative to population size (i.e. number of in-migrants to *Bundesland* plus number of out-migrants divided by population size of *Bundesland*) which ranges from 0.07 to 0.13 (calculated from *Statistisches Bundesamt, 2010b*). Relative in and out migration assesses the degree to which the electoral process may be influenced by socially similar people counting together. We expect that higher levels of migration lead to recruiting more socially heterogeneous groups of election officials who count together and thereby results in more vigilant counting and supervision. Consequently, the sign of the coefficient should be negative.

Third, distance between the largest and the second-largest party (ranging from 1.00 to 27.3 percentage) (calculated from *Bundeswahlleiter, 2010*) depicts the degree of partisan dominance in a *Bundesland*. The more dominant a party is, the more likely the counting process is

biased. The dominant party is in a position to appoint a large number of returning officers. The estimated effect thus should be positive. The three independent variables only correlate up to a level of about  $\pm 0.2$  among each other. These low correlations indicate that the measures capture very different dynamics across observations and that collinearity is not an issue in the subsequent small N regression analysis.

Table 3, first, displays bivariate correlations between the  $\chi^2$  test results for the CDU/CSU, SPD and Left in 2005 from 16 *Länder*, population size, relative in and out migration and partisan dominance. Note that all correlations involving the Left only encompass meager six observations. Due to the size sensitivity of the  $\chi^2$  test, population size correlates highly and positively with the  $\chi^2$  results for the two major parties with 0.86 and 0.71. In and out migration does not correlate at all with any of the party test results, and the little correlation which exists is not in the expected direction. Partisan dominance correlates positively and to a moderate degree with the two major parties'  $\chi^2$  results and negatively for the Left. Overall, the correlations of the Left do not yield any decisive results. The high correlation of 0.74 between the  $\chi^2$  results for the SPD and the CDU/CSU suggests that both parties'  $\chi^2$  might share common predictors, such as *Land*-specific characteristics.

Second, Table 3 shows three pairs of multiple OLS regressions for the CDU/CSU and the SPD. Regressions 1 and 2 only include population size as a covariate, which, of course, has a statistically significant and positive coefficient. For the CDU/CSU results, population size can explain about 75% of the variance; for the SPD results, it is 51% of the variance. Regressions 3 and 4 test the independent effect of relative in and out migration once population size is controlled for. Like in the bivariate correlations, our measure of social cohesion does not capture any systematic variance. Therefore, there is no evidence for the notion that the nature of the *social* recruiting process influences voting irregularities.

As expected, regressions 5 and 6 yield a positive estimated effect for the partisan dominance measure regardless of population size. It is “statistically significant” at the 0.10 level for the SPD model, but also positively correlated in the CDU/CSU model. The more a winning party dominates in a *Bundesland*, the higher are the  $\chi^2$  test results. The estimates imply that an increase in the lead of the first over the second party of one percentage point is associated with an increase of 0.3 and 2.0 of the  $\chi^2$  party test results for the CDU/CSU and the SPD respectively.<sup>7</sup> This finding confirms our expectation that partisan recruitment is important. More dominant

<sup>6</sup> Theoretically, it would be possible to do a joint analysis of all 190 party-year-*Länder* observations. However, there would be obvious multiple dependencies causing potential autocorrelation here, across elections, *Länder* and parties. Exploring them and accounting for them in various models and robustness exercises is beyond the scope of this article.

<sup>7</sup> As an alternative measure, we also tried the success of each party and its squared term as regressor. At low levels of electoral success of each party, the marginal effects are negative and statistically significant, at mean levels they become positive and at the high levels of success they are positive and statistically significant. This implies a u-shaped relationship between electoral success and  $\chi^2$  results. In essence, they imply the same conclusions as the simple dominance measure. Once a party is very strong and dominant,  $\chi^2$  results tend to be higher. The positive impact of the dominance measure can also be found in other election years. In 2002, the analogous models yield significant coefficients of 0.84 (0.37) for the CDU results and 0.79 (0.41) for the SPD results. In the years 1990, 1994 and 1998, 4 out of 6 coefficients are positive, but none reaches “significance” at 0.10 level.

**Table 3**Exploratory bivariate correlations and OLS regression, Chi<sup>2</sup> results CDU/CSU, SPD, Left in 2005.

	Bivariate correlations Chi <sup>2</sup> test results			OLS regressions					
	CDU/CSU	SPD	Left	(1)	(2)	(3)	(4)	(5)	(6)
				CDU/CSU	SPD	CDU/CSU	SPD	CDU/CSU	SPD
CDU/CSU	1	0.74	−0.66						
SPD		1	0.7						
PDS			1						
Population size	0.86	0.71	−0.06	2.346*	4.766*	2.288*	4.807*	2.269*	4.285*
Rel. in/out migration	0.28	0.12	−0.08	[0.36]	[1.26]	[0.38]	[1.33]	[0.36]	[0.98]
Partisan dominance	0.31	0.57	−0.29			68.93 [93.71]	−49.66 [328.84]		
Constant				8.260*	4.161	1.589	8.968	0.318	1.984*
				[2.54]	[8.75]	[9.43]	[33.10]	[0.22]	[0.62]
N	16	16	6	16	16	16	16	16	16
R <sup>2</sup>				0.747	0.507	0.757	0.508	0.781	0.725

Note: standard errors in brackets, \* signifies significance at least at .10 level with a two-tailed test. Significance of the coefficients can be interpreted as an arbitrary measure of the clarity of the effects.

parties are more likely to recruit affiliated returning officers. This recruitment strategy results in a pool of returning officers who are politically alike. This higher occurrence of politically like-minded people might contribute to irregular counting, either as a means of fraud or of human error.

In sum, the exploratory analysis of the 2005 party results illuminates that the degree to which parties dominate in a state varies positively with higher likelihoods of electoral irregularities. Thus, party political infrastructure matters. The analysis confirmed the role of population size but did not lend support to the idea that the extent of movement in and out of a *Bundesland* effects irregularities.

## 5. Conclusions

This article compares the frequencies of numerals in vote counts of German *Bundestag* elections 1990 to 2005 with the expectations according to the Second Digit Benford Law (Mebane, 2006). The analysis aggregates the candidate and party list votes at district level as well as the party list votes at the state level for the three major parties CDU/CSU, SPD and PDS/Left (only for the East). Deviations of the actual frequencies of numerals from the Second Digit Benford Law are interpreted as indicating irregularities – either intentional fraud or unintentional human error – in vote counting.

First, the results reveal minimal evidence of systematic fraud or mismanagement with regard to candidate votes. Across 1500 tests, only four deviant cases are identified. The implication is that there is no evidence for “easy” fraud in Germany, i.e. manipulation at the district level to influence direct seat outcomes by changing a relatively small number of votes. Second, the results for the Western *Länder* with regard to party votes reveal more violations than for the East. Thus, there is little evidence that problems in adapting democratic practices exist in the East. Third, there are, however, 51 violations in the party list votes for the three parties out of 190 tests. These violations vary not only by region, but also by parties and by *Bundesländer* within regions. Fourth, these patterns might indicate that partisan composition of the volunteers involved in the counting

may contribute to electoral irregularities. In particular, the violations for the PDS/Left in 2002, for the CDU/CSU in Baden-Württemberg and Bavaria and for the SPD in North-Rhine-Westphalia allude to the idea that partisan dominance in a *Bundesland* (and even more so in its sub-regions) or aversion to a former regime party among those citizens involved in the counting might produce higher likelihoods of violations of the Second Digit Benford Law.

An empirical exploration of the 2005 results for the party list votes of CDU/CSU and SPD shows that higher test results (i.e. increasing the likelihood of a Benford violation) correlate positively with the distance between the largest and the second-largest party in a *Bundesland*, even when we control for population size. Thus, the more dominant a party is, the more likely it is that the voting returns of the two big parties violate the Benford test. This makes the argument plausible that a distortion occurs when a political party dominates the recruitment process of volunteers and possibly the counting process.

It is beyond the scope of these data to produce a solid foundation for what might be the reasons behind these patterns of electoral irregularities. However, the results warrant deeper investigation into the recruitment process of officials on election day. For example, a survey of all recruitment practices in the 299 districts might reveal some biases (see for a rare example of an article dealing with these issues Buchstein, 2004). Then again, a representative survey of volunteers from several *Bundesländer* might reveal the extent to which there might be evidence of social or political biases at play. Moreover, a close-up study of counting in selected districts, which is public anyway, could be very revealing with a special emphasis on whether counting in socially and politically homogenous groups produces different results than in heterogeneous groups.

## Acknowledgments

Equal authors are listed in alphabetical order. We would like to thank the Max Planck Institute for the Study of Societies Cologne for financing the purchase of the data. We would also

like to thank Regine Spohrer for advising us with the map. Our R and STATA files are available for replication purposes from our websites (<http://individual.utoronto.ca/cbreunig/>, <http://www.achimgoerres.de>). Earlier versions were presented at the 2009 EPOP conference in Glasgow and an internal seminar at the Institute for Empirical Social and Economic Research of the University of Cologne. We would like to thank all participants as well as Bernd Beber, Neil Nevitte, Walter Mebane and the anonymous reviewers for comments.

## Appendix

### *Further details on the counting procedure*

On election day, a Sunday, the precinct officers and their assistant have to arrive at 7 a.m. and keep the station open from 8 a.m. to 6 p.m. with a minimum of two people present at all times, one of whom has to be either the precinct returning officer or his or her replacement. They are equipped with – among other things – the complete electoral register of the precinct, a list of late-registered voters, seals for the ballot box and packing material for the ballots. The precinct officers have to remind all assistants to be impartial and to abide by all rules. Stations are closed at 6 p.m., which is directly followed by the counting procedure. The counting procedure has to end up with results for the number of eligible voters in that precinct, the number of voters, the number of valid and invalid candidate votes and party list votes, the number of votes for each candidate, the number of votes for each party list.

First, all ballot papers have to be unwrapped, counted, and the number compared against the number of electoral passes (*Wahlscheine*) and the list of registered voters. Whenever the counts are not the same, the counts have to be repeated, and problems noted down on official notification paper. Second, the precinct returning officer supervises the assistant counting all ballot papers into four different kinds of piles: (a) one for each party list where the first vote is for the candidate of the party that also received the second vote; (b) one pile with votes for different parties with first and second vote, or where only one vote is valid; (c) one pile for unmarked ballot papers; (d) one pile with unclear markings. Then, the all piles have to be checked by the returning officer and his or her replacement, the results be stated loudly and noted down. Further regulations prescribe how to deal with unclearly marked ballot papers.

At the end of the counting procedure, the results are orally stated in the room, so that everybody involved in the counting procedures gets to know the final results of the precinct. Electronically or by telephone, the results are then transmitted either directly to the district returning officer or to the local administrative district that hosts several precincts. The district returning officer calculates the results for the district by collecting the results from the precincts and from postal voting (that are counted according to another procedure) and takes the result to the state returning officer who then aggregates the state results and transmits them to the federal returning officer. District, state and federal returning officers have to check for the lawfulness of the procedure and indicate any concerns they may have.

At the precinct, the ballot papers have to be packed together in a certain way, and a standardized document be filled out that has to be signed by the precinct returning officers and their assistant. If a person does not want to sign, the refusal and its rationale have to be documented. The precinct returning officer personally and immediately has to transport the material to the local administrative authority. The local authority keeps the ballot papers until these are destroyed once the election results cannot be invoked anymore (usually 3 months) and sends the signed paper work to the district returning officer.

## References

- Alvarez, R.M., Hall, T.E., Hyde, S.D., 2008. *Election Fraud: Detecting and Detering Electoral Manipulation*. Brookings Institution Press, Washington, D.C.
- Beber, B., Scacco, A., 2008. What the Numbers Say: a Digit-Based Test for Election Fraud Using New Data from Nigeria. Columbia University, New York.
- Beber, B., Scacco, A., 2009. The Devil Is in the Digits. *Washington Post*. 20 June 2009.
- Benford, F., 1938. The law of Anomalous numbers. *Proceedings of the American Philosophical Society* 78 (4), 551–572.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery rate: a Practical and Powerful approach to multiple testing. *Journal of Royal Statistical Society, Series B* 57 (1), 289–300.
- Benjamini, Y., Yekutieli, D., 2005. False Discovery rate-Adjusted multiple Confidence Intervals for selected Parameters. *Journal of American Statistical Association* 100 (March), 81.
- Buchstein, H., 2004. Das Verschwinden der armen Wahlhelfer. *Leviathan* 32, 309–318.
- Bundeswahlleiter, 2010. Bundestagswahl 2005 (cited 01–12 2010). Available from: [http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW\\_BUND\\_05/](http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_05/).
- Buttorff, G., 2008. *Detecting Fraud in America's Gilded Age*. University of Iowa.
- Chiang, L., Ordeshook, P.C., 2009. *Fraud, Elections and the American Gene in Taiwan's Democracy*. National Cheng Kung University (unpublished manuscript).
- Deutscher Bundestag, 2000. *Datenhandbuch zur Geschichte des Deutschen Bundestages 1949 bis 1999*. Deutscher Bundestag, Berlin.
- Deutscher Bundestag, 2009a. *Bilanz der Tätigkeit des Ausschusses für Wahlprüfung, Immunität und Geschäftsordnung in der 16. Wahlperiode*. Deutscher Bundestag, Berlin.
- Deutscher Bundestag, 2009b. *Wie wird man Wahlhelfer?* (cited 03-07 2009). Available from: [http://www.bundestag.de/btg\\_wahl/wie/wahlhelfer/index.html](http://www.bundestag.de/btg_wahl/wie/wahlhelfer/index.html).
- Feldkamp, M.F., Ströbel, B., 2005. *Datenhandbuch zur Geschichte des Deutschen Bundestages 1994 bis 2003*. Nomos, Baden-Baden.
- Freiberger, M., 2009. Does the Iranian Election Stand Up to Statistics? *Millenium Mathematics Project*, University of Cambridge, 07-07-2009 2009 (cited 23-11 2009).
- Gelman, A., 2009. Unconvincing (to me) use of Benford's law to demonstrate election fraud in Iran, 17-06-2009 (cited 03-12 2009). Available from: <http://www.stat.columbia.edu/gelman/blog/>.
- Gschwend, T., 2007. Ticket-splitting and strategic voting under mixed electoral rules: evidence from Germany. *European Journal of Political Research* 46, 1–23.
- Hill, T.P., 1995. A statistical derivation of the significant-digit law. *Statistical Science* 10 (4), 354–363.
- Hill, T.P., 1998. The first digit Phenomenon. *American Scientist* 86, 358–363.
- Landesregierung Berlin, 2009. *Verwaltungsvorschriften über den Ausgleich für die Tätigkeit von Dienstkräften der Berliner Verwaltung in Wahl- und Abstimmungsvorständen bei den allgemeinen Wahlen und Abstimmungen*. *Amtsblatt für Berlin* 13/2009, 739.
- Lehoucq, F., 2003. Electoral fraud: causes, types, and Consequences. *Annual Review of Political Science* 6 (1), 233–256.
- McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a Feather: homophily in social Networks. *Annual Review of Sociology* 27, 415–444.
- Mebane, W.R., 2006. *Election forensics: vote counts and Benford's law*. Political Methodology Society. University of California at Davis.
- Mebane, W.R., 2007. *Statistics for digits*. Summer Meeting of the Political Methodology Society. Pennsylvania State University.
- Mebane, W.R., 2008. Election forensics: the second-digit Benford's law test and recent American elections. In: Alvarez, R.M., Hall, T.E.,

- Hyde, S.D. (Eds.), *Election Fraud: Detecting and Deterring Electoral Manipulation*. Brookings Institution Press, Washington, D.C.
- Mebane, W.R., 2009. Note on the presidential election in Iran (29-06-2009). Available from: <http://www-personal.umich.edu/wmebane/note29jun2009.pdf> (cited 03-12 2009).
- Myagkov, M., Ordeshook, P.C., Shakin, D., 2009. *The Forensics of Election Fraud*. Cambridge University Press, Cambridge.
- Newcomb, S., 1881. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics* 4, 39–40.
- OSCE, 2009. *Federal Republic of Germany: Elections to the Federal Parliament (Bundestag)*. OSCE/ODIHR, Warsaw.
- Roukema, B.F., 2009. Benford's Law Anomalies in the 2009 Iranian Election. *Torun Centre for Astronomy: Nicolaus Copernicus University*.
- Schulz, S., 2009. PRO RELI: 2500 Lehrer Müssen Beim Volksentscheid Helfen. *Berliner Morgenpost*. 06-03-2009.
- Statistisches Bundesamt, 2009. *Statistisches Jahrbuch 2009*. Statistisches Bundesamt, Wiesbaden.
- Statistisches Bundesamt, 2010a. *Bevölkerungsstand: Bevölkerung nach Geschlecht* (cited 01-12 2010). Available from: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Statistiken/Bevoelkerung/Bevoelkerung.psml>.
- Statistisches Bundesamt, 2010b. *Zu- und Fortzüge nach Altersgruppen (über Gemeindegrenzen)* (cited 01-12 2010). Available from: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Statistiken/Bevoelkerung/Wanderungen/Wanderungen.psml>.
- Süddeutsche Zeitung, 2006. Dachauer Wahlfälscher müssen mehr als 116.000 Euro zahlen 08-08-2006.
- Wand, J.N., Shotts, K.W., Sekhon, J.S., Mebane, W.R., Herron, M.C., Brady, H.E., 2001. The Butterfly did it: the Aberrant vote for Buchanan in Palm Beach County, Florida. *American Political Science Review* 95 (4), 793–810.
- Ziblatt, D., 2009. Shaping democratic practice and the causes of electoral fraud: the case of Nineteenth century Germany. *American Political Science Review* 103 (1), 1–21.