Identification, Characteristics and Impact of Faked and Fraudulent Interviews in Surveys

Christin Schäfer[1], Jörg-Peter Schräpler[2,3] and Klaus-Robert Müller[1,4]

[1]Fraunhofer FIRST.IDA, Kekuléstr. 7, 12489 Berlin, Germany
[2]Ruhr University Bochum, 44780 Bochum, Germany
[3]DIW Berlin, Königin-Luise-Str. 5, 14191 Berlin, Germany
[4]University of Potsdam, August-Bebel-Str. 89, 14482 Potsdam, Germany

{christin,klaus}@first.fraunhofer.de
joerg-peter.schraepler@ruhr-uni-bochum.de

June 26th 2004

Abstract

This paper presents two new tools for the identification of faking interviewers in surveys. One method is based on Benford's Law, the other exploits the empirical observation that fakers most often produces answers with less variability than could be expected from the whole survey. We focus on fabricated data in the German Socio-Economic Panel (SOEP). For several samples the resulting rankings of the interviewers with respect to their cheating behavior are given. We further investigate the impact of evident fakers to studies that rely on the data of the survey. If the faked interviews are removed, the results can change dramatically.

1 Introduction

1.1 Faking

In any survey in which the data are collected by personal interviews there is a danger of cheating by interviewers, or that some interviewers may fabricate data. We can distinguish several forms of cheating:

1. First, the most blatant form is when an interviewer fabricates all 'responses' for an entire questionnaire. The U.S. Bureau of the Census refers to this practice as "falsification" or 'fabrication'. Sometimes this practice is also called 'curbstoning', thus named because a census taker "stands at the curb" and guesses the number of residents in a building or house without ever entering (Moore and Marquis 1996). Falsification also include the acceptance of proxy information when self-response is required and the unauthorized use of the telephone when a personal visit is required.

2. A second more subtle form is, when an interviewer asks some questions in an interview and fabricates the responses to others.

3. A third form of cheating is when an interviewer knowingly deviates from prescribed interviewing procedures, such as conducting an interview with someone who is easily reachable and willing to participate in the place of the appropriate person.

In this paper we address the first form of cheating, the fabrication of an entire interview. The paper is organized as follows: in the next section we review previous results on cheating behavior in the literature. In section 1.3 we compare in a first analysis the cheating behavior in the Socio-Economic Panel (SOEP) with the methods described in the literature and compare the results. In section 2 we describe two unconventional approaches for the identification of fakers that uses only the data of the survey without any additional information. The first is based on Benford's Law (cf. section 2.1), the second has the name variability method (cf. section 2.2). The results of the two new methods are given in section 3. In section 4 we study the possible influence of fakers to investigations based on the survey. We finalize with a discussion in section 5.

1.2 Previous results on cheating behavior

As compared to other methodological topics, there are only few studies dealing with cheating by interviewers that have appeared in the literature.

- Crespi (1945) investigate which factors may contribute to cheating behavior. He distinguished between factors relating to questionnaire characteristics (design and length, difficult and antagonistic questions), administrative demoralizers (inadequate remuneration and training of the interviewer) as well as external factors (bad weather, bad neighborhoods, etc.). He proposed a dual strategy of eliminating demoralizers. Furthermore he uses a verification method to deter cheating. Some more recent studies refer to these verification methods and deal with optimal designs of quality control samples to detect interviewer cheating (Biemer and Stokes 1989) and the evaluation of the quality control procedures for interviewers (Stokes and Jones 1989).

- Because of the lack of factual information concerning the nature of interviewer falsification the U.S. Census Bureau implemented an 'Interviewer Falsification Study' in the year 1982 (Schreiner, Pennie, and Newbrough 1988). In this study data were accumulated from fifteen surveys conducted by twelve U.S. Census Bureau regional offices over a five-year period. They found 205 cases of confirmed falsification. Most of these (74%) were detected through reinterview's and the majority (79%) was determined to have fabricated interviews. Their results provide evidence that the shorter the length of service, the more likely an interviewer will falsify data (Schreiner, Pennie, and Newbrough 1988). Furthermore, when new interviewers falsify data, it is usually a relatively high proportion of their assignments and they tend to fabricate entire interviews. Interviewers with five or more years of experience usually falsify a smaller proportion of their assignments and tend to classify eligible units as ineligible (Hood and Bushery 1997).

- Other studies deal with the 'quality' of faked interviews and the impact of fabricated data on substantive analysis. Reuband (1990) shows that students are able to reproduce data in fictive interviews using given demographic variables of real respondents.

- Schnell (1991) performed a study in which he substituted 220 real interviews of the German General Social Survey (ALLBUS 1988, N = 3052) with fictive interviews fabricated by sociology students and university colleagues. He analyzes the quality of the fabricated data and the robustness of substantive empirical results by comparing the German General Social Survey with the substituted faked data. His main result is that univariate statistics like proportions, means and variances are relatively robust against typical amounts of fabricated data in surveys (less than 5%). Nevertheless he also found some minor effects on multivariate statistics like multiple regressions. Moreover, using simulations he shows that higher proportions of fabricated data in surveys will have a serious impact on multivariate statistics and data quality.

- In the ALLBUS 1994 the ADM design was replaced with a new sampling design, which offers the opportunity to systematically check that the interviews (N = 3505) are performed correctly. The interviewers are given the address and the names of the respondent directly. In six percent of the cases irregularities were detected; half of them turned out to be faked by the interviewers (Koch 1995). These fabricated data (n = 45) are found after the routine monitoring by the data collection institute via the postcard method, which detected fifteen faked interviews in this survey. Another finding was that interviewers who cheat are mainly younger persons with higher education (Abitur) and with a relatively high workload (number of interviews).

- A rare debacle caused by faked interviews is mentioned by Diekmann (2002). In the German city Rostock a traffic study about drivers was carried out by means of 600 face-to-face interviews. Eighty cases were later recontacted for another study, which showed that sixteen of the former interviews were completely or partly fabricated by the interviewer. If we extrapolate this to the whole sample, that amounts to a share of 20% fakes.

In the next section we analyze the cheating behavior in the SOEP with procedures proposed in the literature mentioned above.

1.3 Fabrication within the Socio-Economic Panel

We focus on fabricated data in the German Socio-Economic Panel (SOEP). From the fieldwork organization we get faked records. Notice, that other fieldwork organizations hide this problem. Furthermore we get some hints about the quality control procedures which are performed as standard to detect fakes. These verification methods as well as ‘conventional’ statistical tests of stability and consistence are the ones proposed by Crespi (1945) (see section 1.2).

Descriptive analysis of cheating behavior. The SOEP consists of several samples. Fabricated data were always found in the first wave of each sample (with the exception of the East German sample C and the small sample D). Nevertheless, one interviewer was able to fabricate data for the first two waves of sample E without raising suspicion until wave 3. Table 1 shows the (detected) amount of fabricated data. The first wave of samples A and B contains only 0.6 and 1.5% fabricated data, respectively, and the first wave of sample E contains about 2% faked household interviews. In the following wave approximately 1% of fabricated data was identified in sample E. In the first wave of sample F1 the cheating was lowest: only 0.1% of the interviews were detected as fabricated.

Table 1: *Proportion of detected fabricate data in the SOEP*

	Household interview			*Personal interview*		
Sample	*Valid cases*	*Fabricated cases*	*% of fakes*	*Valid cases*	*Fabricated cases*	*% of fakes*
1984						
Sample A	4528	26	0.6	9115	59	0.6
Sample B	1393	22	1.5	3175	45	1.4
1998						
Sample E	1056	23	2.1	1910	47	2.4
1999						
Sample E	886	11	1.2	1629	22	1.3
2000						
Sample F1	5848	8	0.1	10470	11	0.1
Total	13711	48	0.4	26299	184	0.7

Source: SOEP 1984 - 2000 (own calculation)

Quality control and detecting cheating interviewers in the SOEP. In contrast to cross-sectional surveys, curbstoning is extremely difficult in complex long-term panel studies like the SOEP (German Socio-Economic Panel Study) because the respondent is interviewed face to face every year, and because a consistency check between waves shows irregularities immediately. Hence we can assume that fabricated data will be a problem mainly in the first wave and will be detected quickly after conducting the second wave. As shown above this is clearly the case in the SOEP.

The most common method used for detecting interviewer cheating in face-to-face surveys is the verification method where a sample of an interviewer's assignment is recontacted in order to verify that an interview was conducted. In this sense the SOEP provides a unique opportunity to identify fabricated data. Falsifications are detected in several ways:

1. Most fakes are identified easily by comparing data of two waves. If data deviate considerably from the data of the previous year(s), the interview control department contacts the respective households by phone and the household members are asked to verify the data. If there is a change in interviewer in the following wave, in the case of falsified data the new interviewer cannot confirm the composition of the household as recorded in the address protocol.

2. After the interview, all respondents receive a "thank-you" letter and a small gift by mail for having given the interview. Hence, if the interview did not take place, the intended respondent is likely to contact the fieldwork organization, which then becomes aware of the falsified interviews. If the recorded address does not exist, the interviewer control department is informed.

3. Due to problems with curbstoners in sample E of SOEP (1998), for sample F (2000) all households were recontacted after interviewing and asked to verify the household composition.

Additional control routines are implemented to further secure the quality of the fieldwork. The fieldwork organization uses mainly experienced interviewers for the SOEP project. The average length of service in the first wave is approximately five years.

Area characteristics for detected fabricated data in the SOEP. Because Biemer and Stokes (1989, p.25) find that in the two large demographic surveys cheating behavior differed between urban and rural areas we examine these kind of differences. In the United States, the CPS and the NCS, 87% of the falsified interviews came from urban areas, and only 13% of the falsified interviews coming from rural areas. Since 70% of the sample is located in urban areas for these surveys, there is evidence of a higher degree of cheating in urban areas. Table 2 shows the frequency of fabricated household interviews in Sample A, B, and E of the SOEP by number of residents in the area. The proportion of falsification in cities ($\geq 100,000$ residents) is in sample A/B 52.1%, and the proportion of cities in the non-faked data of sample A/B is only 40.4%. Also the proportion of rural areas is in the faked sample higher than in the non-faked sample. These differences are statistically significant on a 1% level (Chi-Square = 1452). Nevertheless we find no statistically significant area effect in sample E if we only differentiate between cities and non-cities (Chi-square = 0.06). This finding suggest that there may be only an unsystematically area effect. This may be important because the unknown true data which are fabricated by the cheating interviewers have the same area distribution as the faked interviews. Systematic differences of faked and non-faked interviews in the area characteristics would suggest that the distributions of the unknown true data are different from the known non-faked data.

TABLE 2: *Distribution of fakes by area characteristics*

No Residents in Area	A+B		A+B		E		E		total		total	
	faked		non-faked		faked		non-faked		faked		non-faked	
	no	%	no	%	no	%	no	%	no	%	no	%
$> 100,000$	25	52.1	2369	40.4	12	35.3	375	35.5	37	45.1	2744	39.6
$20 - 100,000$	1	2.1	1543	26.3	22	64.7	240	22.7	23	28.0	1783	25.7
$< 20,000$	22	45.8	1959	33.4	-	0.0	441	41.8	22	26.9	2400	34.7
Overall	48	100	5871	100	34	100	1056	100	82	100	6927	100

Source: SOEP Sample A, B, and E (own calculation)

Characteristics of cheating interviewers. Only very little is known about the characteristics of interviewers who cheat in surveys. Koch (1995, 97) shows that younger interviewers with a higher education level have more inconsistencies in their interviews than others. Table 3 lists some characteristics. All interviewers who fabricated data (N = 9) are middle-aged males. We find no education effects; cheating interviewers may have a university degree or only a primary school education. In addition in sample A cheating interviewers have on average a higher assignment of household interviews (18.3) than the interviewers in the non-faked data (9.6). In sample E the difference between the average assignments (non-faked data: 7.32; faked data: 11.67) is neither statistically significant on a 5% nor on a 10% level. In the first wave of all samples, almost all cheating interviewers falsified their entire assignments, and one interviewer in samples A and B falsified just one of over 43 personal interviews. Whereas the first interviewers in sample A/B are more than three years in service, the latter one with only one faked interview works only two years for the fieldwork organization. But each of them was working on this panel study for the first time. We can assume that they were not aware of the effectiveness of the quality control in this panel study and of the fact that fakes in this design are easily identifiable. This finding is not in line with results of Hood/Bushery (1997). They have found that cheating interviewer who falsify a relatively high proportion of their assignments are inexperienced interviewers.

Table 3: *Characteristics of interviewers with fabricated data*

					occup.	*Number*	*fabric.*	*Number*	*fabric.*	*years in*
	Intnr	*gender*	*birth*	*educ.*	*main job*	*hh-int.*	*hh-int.*	*pers. int.*	*pers. int.*	*Service*
Sample A + B										
1984	43800	male	1942	univ.	part-time	14	12	38	36	16
1984	128279	male	1935	sec. sch.	full-time	18	18	35	35	5
1984	139378	male	1928	sec. sch.	full-time	17	17	32	32	4
1984	165824	male	1952	univ.	part-time	24	1	43	1	2
Sample E										
1998	236837	male	1966	univ.	student	22	12	33	25	4
1998	249281	n.k.	n.k.	n.k.	n.k.	1	1	2	2	n.k.
1998	238037	n.k.	n.k.	n.k.	n.k.	1	1	2	2	n.k.
1998	236807	n.k.	n.k.	n.k.	n.k.	1	1	2	2	n.k.
1998	249289	male	1951	prim. sch.	full-time	12	10	25	20	1
1999	249289	male	1951	prim. sch.	full-time	11	11	22	22	2
Sample F										
2000	270857	male	1949	sec. sch.	full-time	8	8	11	11	¡ 1

Source: SOEP - Interviewer data set 1984 - 2000 (own calculation)

2 Two new methods for fraud detection in Surveys

2.1 Benford's Law

Benford's Law is an empirical ”law” which states that in many tables of numerical data, the significant digits are not uniformly distributed as might be expected, but rather obey a certain logarithmic probability distribution (Hill 1996b). According to Hill (1999) in 1881, the astronomer Newcomb (Newcomb 1881) explained that his discovery of the significant-digit law was motivated by an observation that the pages of a book of logarithms were dirtiest in the beginning and progressively cleaner throughout. Nevertheless the law is named for Dr. Frank Benford, a physicist who had made the same observation. In 1938 he embarked on a mathematical analysis of 20,229 sets of numbers, including such wildly disparate categories as the areas of rivers, baseball statistics, numbers in magazine articles and street addresses (see table 4, Benford 1938).

TABLE 4: *The distribution of leading digits in Benford's data sets in percentages (Benford 1938)*

Group	Title	1	2	3	4	5	6	7	8	9	Count
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P.Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Weight	27.7	25.3	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	n^{-1}, $\sqrt{n}$,...	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	Gigest	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2, \ldots, n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
	Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
	Predicted	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6	

He found that all these seemingly unrelated sets of numbers followed the same first-digit probability pattern. In most cases the number 1 turned up as the first digit about 30 percent of the time, more often than any other. Benford derived a formula to predict the frequency of numbers found in many categories of statistics. The leading significant (non-zero) digit obeys the law

$$Prob(\text{first significant digit} = d) = log_{10}\left(1 + \frac{1}{d}\right), \qquad d = 1, 2, \ldots, 9$$

Hence, a number chosen at random has leading significant digit $d = 1$ with probability 0.301, a leading digit $d = 2$ with probability 0.176 and so on monotonically down to probability 0.046 for leading digit $d = 9$. The general law for second and higher significant digits and their joint distribution is (Hill 1996a, 1999):

$$Prob(D_1 = d_1, \ldots, D_k = d_k) = log_{10}\left[1 + \left(\sum_{i=1}^{k} d_i \times 10^{k-i}\right)^{-1}\right] \tag{1}$$

where $d_1 \in \{1, 2, \ldots, 9\}$ and $d_j \in \{0, 1, 2, \ldots, 9\}, j = 2, \ldots, k$. Therefore the joint probability $Prob(D_1 = 1, D_2 = 5, D_3 = 2) = log_{10}(1 + (152)^{-1} \approx 0.0028$.

From equation 1 follows that the significant digits are dependent and not independent. Table 5 shows the joint distribution for the first two digits. It can easily be seen that the joint probability that the second digit is 3, given that the first digit is 1, is $P(D_1 = 1, D_2 = 3) \approx 0.0299$, but $P(D_1 = 1) \cdot P(D_2 = 3) \approx 0.0314$.

This dependence among significant digits decreases rapidly as the distances between the digits increases. The table below table 5 shows the distribution of the first till fourth significant digits. We can recognize that the distribution of the nth significant digit approaches the uniform distribution on $0, 1, \ldots, 9$ exponentially fast as $n \to \infty$ (c.f. Hill 1995, p.355).

For many years the status of this law was little more than a numerical curiosity but practical implications began emerge in the 1960s (Scott/Fasli 2001). It was recognized that the suggestion that almost 1/3 of the numbers processed began with digit "1" could have implications for the design of efficient computers (Hamming 1970; Knuth 1981). In recent years it has been successfully used to detect fraudulent financial data (Nigrini 1999).

Despite this rather slender empirical support (Scott/Fasli 2001), there is disagreement about whether this law is a necessary mathematical truth or a contingent property of nature.

Table 5: *The joint distribution for the first two digits in according with Benford*

	D_2										
D_1	0	1	2	3	4	5	6	7	8	9	$\sum_{D_2=0}^{D_2=9}$
1	0.0413927	0.0377886	0.0347621	0.0321847	0.0299632	0.0280287	0.0263289	0.0248236	0.0234811	0.0222764	0.3010300
2	0.0211893	0.0202034	0.0193052	0.0184834	0.0177288	0.0170333	0.0163904	0.0157943	0.0152400	0.0147233	0.1760913
3	0.0142404	0.0137883	0.0133640	0.0129650	0.0125891	0.0122345	0.0118992	0.0115819	0.0112810	0.0109954	0.1249387
4	0.0107239	0.0104654	0.0102192	0.0099842	0.0097598	0.0095453	0.0093400	0.0091434	0.0089548	0.0087739	0.0969100
5	0.0086002	0.0084332	0.0082725	0.0081179	0.0079689	0.0078253	0.0076868	0.0075531	0.0074240	0.0072992	0.0791812
6	0.0071786	0.0070619	0.0069489	0.0068394	0.0067334	0.0066306	0.0065309	0.0064341	0.0063402	0.0062489	0.0669468
7	0.0061603	0.0060741	0.0059904	0.0059089	0.0058295	0.0057523	0.0056771	0.0056039	0.0055325	0.0054629	0.0579919
8	0.0053950	0.0053288	0.0052642	0.0052012	0.0051396	0.0050795	0.0050208	0.0049634	0.0049073	0.0048525	0.0511525
9	0.0047989	0.0047464	0.0046951	0.0046449	0.0045958	0.0045476	0.0045005	0.0044543	0.0044091	0.0043648	0.0457575
$\sum_{D_1=1}^{D_1=9}$	0.1196793	0.1138901	0.1088215	0.1043296	0.1003082	0.0966772	0.0933747	0.0903520	0.0875701	0.0849974	

Source: own calculation

Marginal distribution for first - fourth significant digit d

d	$P(D_1 = d)$	$P(D_2 = d)$	$P(D_3 = d)$	$P(D_4 = d)$	uniform distribution
0	-	0.1196793	0.1017843	0.1001761	0.1000000
1	0.3010300	0.1138901	0.1013759	0.1001368	0.1000000
2	0.1760913	0.1088215	0.1009721	0.1000976	0.1000000
3	0.1249387	0.1043296	0.1005729	0.1000584	0.1000000
4	0.0969100	0.1003082	0.1001780	0.1000193	0.1000000
5	0.0791812	0.0966772	0.0997870	0.0999802	0.1000000
6	0.0669468	0.0933747	0.0994013	0.0999412	0.1000000
7	0.0579919	0.0903520	0.0990192	0.0999022	0.1000000
8	0.0511525	0.0875701	0.0986411	0.0998632	0.1000000
9	0.0457575	0.0849974	0.0982671	0.0998243	0.1000000

Source: own calculation

2.1.1 Explanations of Benford's Law

There is empirical evidence that many classes of true data sets follow Benford's Law. It has been found to apply to many sets of financial data, including income tax and stock exchange data, corporate disbursements and sales figures, demographics and scientific data (e.g. Nigrini 1999), as well as numbers gleaned from newspaper articles (Benford 1938; Hill 1999). Stock prices may seem to be a single distribution, but their value actually stems from many measurements (salaries, the cost of raw material and labour) and so it is expected that they will follow Benford's Law in the long run. A recent study about whether tax returns in Germany follows Benford's Law shows that not all but most parts conform to the logarithmic distribution.

On the other hand, truly random numbers do not confirm to Benford's Law, because the proportion of leading digits in such numbers are, by definition, equal. Those data sets most likely follow to Benford's Law have numbers which do not contain a built-in maximum and describe the sizes of similar phenomena (Nigrini 1999). Assigned numbers, such as Social Security numbers or bank accounts, will not conform to it. Furthermore deviations from the law's prediction can be caused by merely rounding numbers. Moreover, the sample of numbers should be large enough to give the predicted proportions a chance to assert themselves (Pinkham 1961) and the sets of numbers should essentially be subsets of a larger series and not just huge chunks of such series.

The random-samples-from-random-distribution Theorem from Hill (1995). A plausible theoretical explanation for the appearance of this logarithmic distribution is the *random-samples-from-random-distribution theorem* by the mathematician Hill (1995). He shows "that if probability distributions are selected at random, and random samples are then taken from each of these distributions in any way so that the overall process is scale (or base) neutral, then the significant digit frequency of the combined sample will converge to the logarithmic distribution." (Hill 1995, 360). It is not required, that individual realizations of a random variable be scale or base invariant. But it is necessary that the sampling process on the average does not favor one scale over another (Hill 1995, p.361).

This theorem is important for the answer to the question if Benford's Law is feasible for survey data because survey data contain different variables with different distributions. Therefore we can test if the chosen mixture of variables from survey data are scale unbiased and if this is the case it is reasonable that this mixture of data follows Benford's Law.

Summary of the necessary requirements. Although we have found empirical evidence for the validity of Benford's Law in the literature there is also evidence, that many natural data sets don't confirm to this logarithmic distribution (c.f. Scott/Fasli 2001). Hence it is important to summarize all necessary requirements of the data characteristics for the usage of Benford's distribution to detect fraudulent data. Some of these requirements derive from simulation results (Scott/Fasli 2001), others are findings from practical applications (Nigrini 1999) or theoretical analysis (Hill 1995).

- The data set should not contain a built-in-maximum because the frequency of these values will occur in the digit analysis more often and will cause biased results (Nigrini 1999).

- The data set should not contain assigned numbers like social security numbers or bank accounts (Nigrini 1999).
- The data set should have only positive values with an unimodal distribution whose modal is not zero (Scott/Fasli 2001).
- The data set should have a positive skewed distribution in which the median is no more than half of the mean. Hence the data set should contain more smaller than larger values.
- The data set should not emanate from statistical procedures like calculated means or variances that emanate from other data (Mochty 2001).

Furthermore another requirement is the sufficient sample size of the data set. The larger the sample size the better should be the fit to Benford's distribution if all of the above requirements are satisfied.

We can now answer the question whether or not Benford's Law can be used to identify fabricated data in surveys. Unlike financial data, many variables in these survey databases are dichotomous or categorical (like gender, marital status and occupation) or are assigned numbers like household numbers. Hence we have to restrict our Benford analysis on continuously data like monetary variables.

2.2 Variability Method

The variability method is based on the empirical finding that the variance of all answers across all questionnaires delivered by a faking interviewer is lower than the variance that is achieved by questionnaires of non-fabricated interviews. There are several points that could explain the absence of variance in fabricated interviews:

- Fakers tend to answer every question. Thus they produce less missing values.
- In questions where one needs to assign a score, like (1) 'I agree' up to (5) 'I disagree', fakers tend to make a cross in the middle. Extreme values are avoided.
- Since the interviewers know the questionnaire and understand the meaning of the questions by faking they will not produce any astonishing answers. Such answers can be found in non-fabricated interviews because the interviewees have misunderstood a question.

This list is not complete.

The variability method consists of the following steps: first measure the variance inside of all questionnaires of an interviewer, second compare this value to the expected variance for a questionnaire cluster of the given size on the whole survey. More formally, let I_i, $i = 1, \ldots, n$, denote the interviewer i, and n is the number of interviewers that have conducted the survey. The number of questionnaires Q_j is given by m with $j = 1, \ldots, m$ and $m = m_1 + \ldots + m_i$, where m_i denotes the number of questionnaires delivered by interviewer I_i. Without taking into account any meaning of the answers – whether a 5 encodes for '5 years' or for 'I disagree' – we calculate the variance for every question $Q(k)$, $k = 1, \ldots, l$ on all questionnaires Q_j of an interviewer I_i and sum up over all questions:

$$T_{I_i} = \sum_{k=1}^{l} \sum_{j=1}^{m_i} (Q_j(k) - \overline{Q(k)})^2. \qquad (2)$$

Here, $\overline{Q(k)}$ denotes the mean for question $Q(k)$ and the index j accounts all questionnaires Q_j, $j = m_{i1}, \ldots, m_{im_i}$ of the interviewer I_i.

The distribution of the test statistic T is estimated using a resampling approach on the whole survey. From this distribution we can derive a probability of the observed value. In the following we will denote this probability with plausibility. By sorting the interviewers with respect to the plausibility they achieved we obtain an interviewer ranking. The interviewers with lowest plausibility are at the top of the ranking. They are considered to be potential fakers.

3 Empirical Results

3.1 Results with Benford's Law

3.1.1 Description of the data used for Benford's Law

First we have to give some descriptive measurements about the data we will use for examining with Benford's Law. The selected data are restricted to variables with monetary values. Besides the monthly gross- and net-income the data sets contain variables like gross amount of Christmas or vacation bonus, gross amount of monthly unemployment benefits or monthly subsistence allowance, gross amount of early retirement benefits, amount of taxes as well as many other monetary variables. Figure 1 shows the estimated distribution for the first wave of sample A/B, the first two waves of Sample E and the first wave of the Sample F. The wave specific figures contain also the number of variables in the wave specific data set, the number of values in this data set (N), the average mean, the standard deviation and the median.

We can recognize that all distributions have rather the same shape, the distributions are unimodal and the medians are always lower than the means and led to positive skewed distributions. An unimodal positive skewed distribution is one important requirement for the use of Benford's Law (Scott/Falsi 2001).

3.1.2 Overall fit to Benford's distribution

In this section we examine the overall goodness of fit, to ascertain that we can use the logarithmic distribution for detecting conspicuous interviewer clusters. If the overall digit distribution in each wave doesn't follow roughly Benford's distribution, we can not abide that this will be the case in interviewer clusters. The following figures 2 to 5 show the first digit and the first two digit distribution of the selected data sets. The figures show rather similar distributions, the shapes of the first digit distributions are rather close to Benford. The distributions for the first two digits in figures 2 - 5

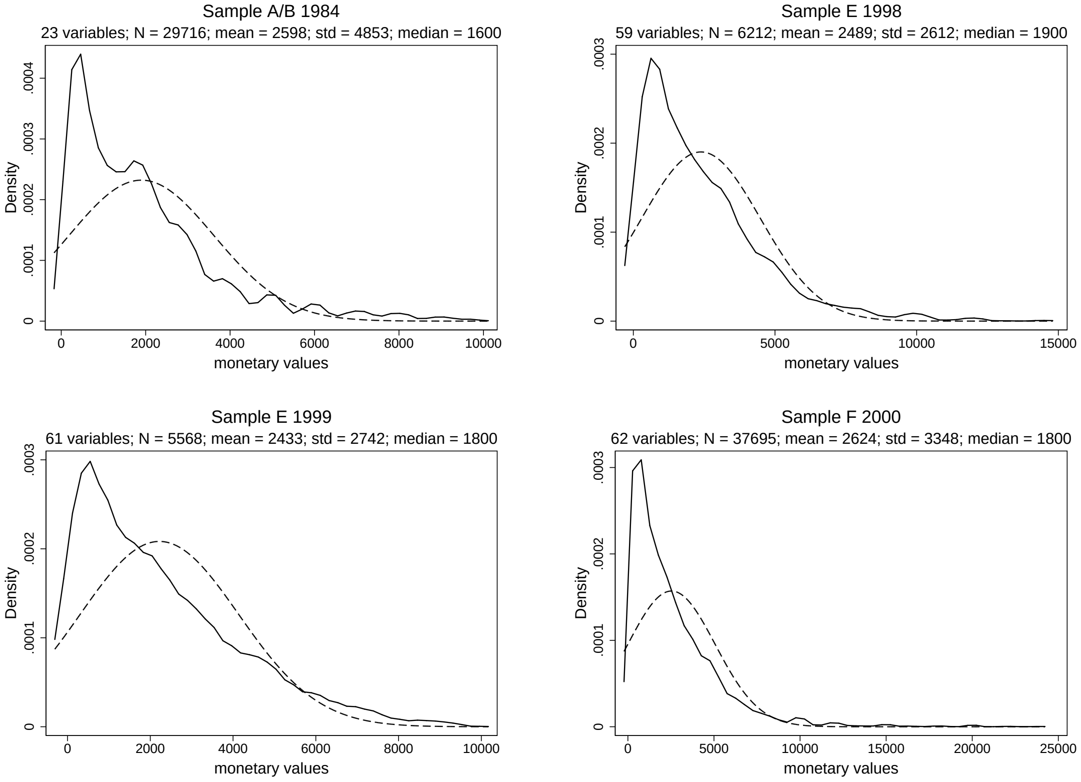

FIGURE 1: *Kernel density estimation for the distribution of the selected monetary data sets of Sample A/B, Sample E and Sample F*

show especially higher proportions for numbers like $10, 20, 30, \ldots, 90$, that are a result of respondent's rounding behavior.

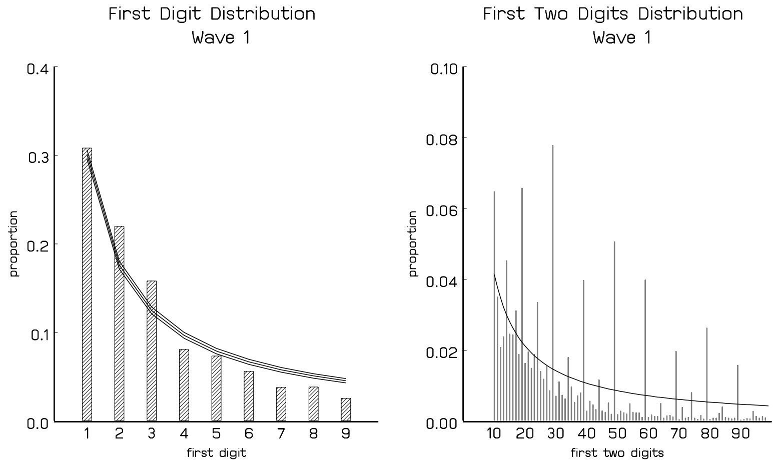

FIGURE 2: *Sample A/B Wave 1*, $\chi^2 = 1279$, *for first digit distribution*, $N = 29,716$

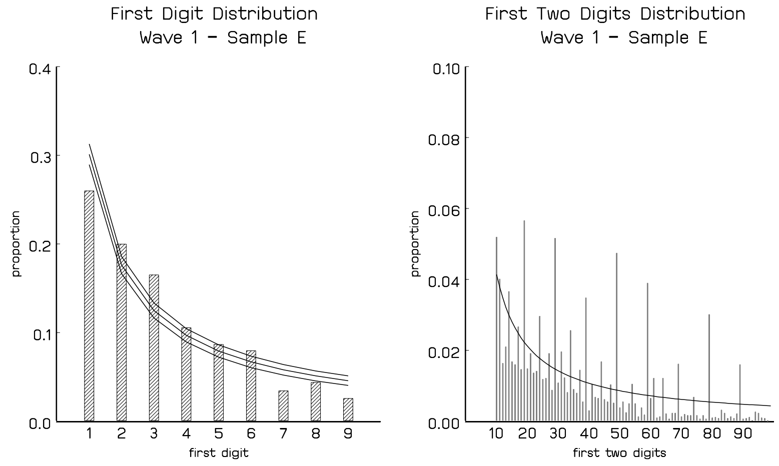

FIGURE 3: *Sample E Wave 1*, $\chi^2 = 272$, *for first digit distribution*, $N = 6,212$

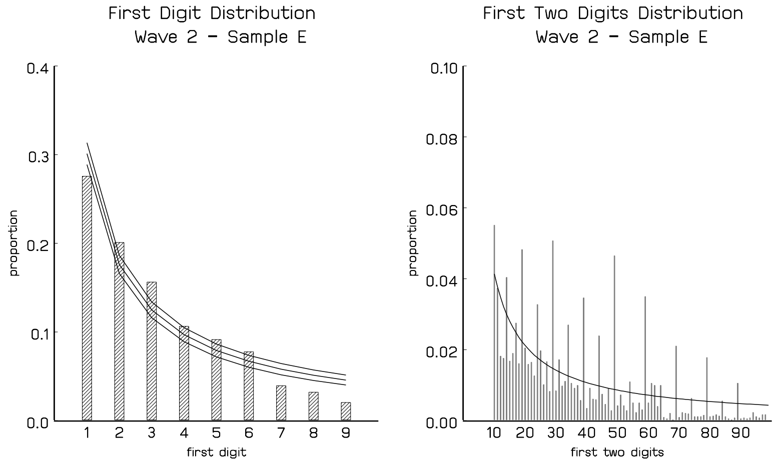

FIGURE 4: *Sample E Wave 2*, $\chi^2 = 246$, *for first digit distribution*, $N = 5568$

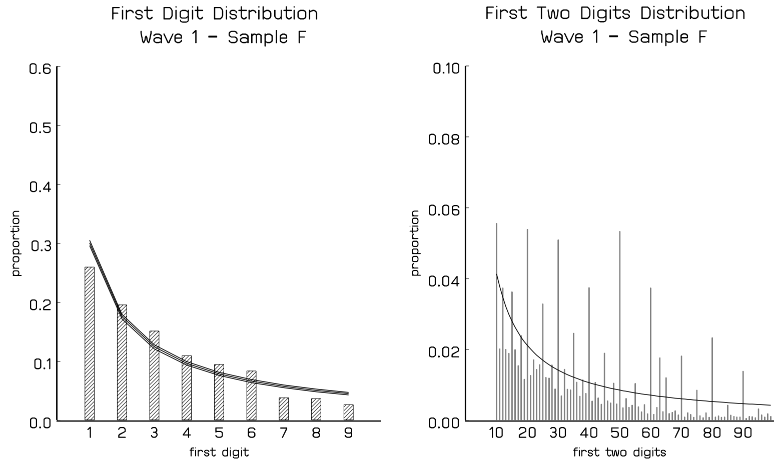

FIGURE 5: *Sample F Wave 1*, $\chi^2 = 1,550$, *for first digit distribution*, $N = 37,656$

3.1.3 Interviewer-ranking with Benford

Now we are going to check whether it is possible to detect cases with fabricated data using Benford's Law. We have shown that the interviewers fabricate a large proportion of their assignment. Therefore it gives more statistical power if we analyze whole clusters of interviews per interviewer (”interviewer cluster”) rather than single questionnaires. If real survey data follows the logarithmic distribution and fabricated survey data not, we should be able to identify these clusters of fabricated interviews and to test them for significance.

To explore the fit of each cluster we calculate chi-square values

$$\chi_i^2 = n_i \sum_{d=1}^{9} \frac{(h_{d_i} - h_{b_d})^2}{h_{b_d}}$$

where n_i is the number of first digits in the interviewer cluster i, h_{d_i} is the observed proportion of digit $d = 1, \ldots, 9$ in interviewer cluster i and h_{b_d} is the proportion of digit d under Benford's distribution.

The usage of chi-square values has the disadvantage that the values depend also partly on the number of observations with the consequence that we will get higher chi-square values for larger interviewer clusters with many digits. A better solution might be the calculation of probabilities for the realized chi-square values on the basis of a resampling method like a bootstrap.

An approximation of the probability of obtaining a value of test statistic (chi-square values) more extreme than that actually observed $Prob(\theta > \hat{\theta})$ can be obtained directly from the proportion of bootstrap replications B higher than the original estimate $\hat{\theta}$

$$P(perc) = Prob(\theta > \hat{\theta}) = 1 - (\frac{\#\hat{\theta}^*(b) < \hat{\theta}}{B}) \qquad (3)$$

These probabilities reflect the *plausibility* of the fit to Benford independent of the number of digits in the cluster. Our hypothesis is that cheating interviewers will have very low probabilities. Hence it might be useful to construct interviewer-rankings by probability values.

Table 6 shows the ranking for the samples with known and already detected fakes. The faking interviewer are marked bold. We could recognize that several cheating interviewers occurs on the top of the list because their fit statistics are not plausible. If we regard the first ten interviewers as suspicious, we identify with Benford in sample A one of three faker, in sample E wave 1, three of five fakers, in wave 2 one of one and in sample F also one of one faker.

Table 6: *Interviewer-ranking with Benford (faking interviewer bold)*

	Sample A/B, wave 1					Sample E, wave 1					Sample E, wave 2					Sample F, wave 1			
Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)
1	**128279**	**122**	**52.30**	**0.0020**	**1**	**236837**	**221**	**49.07**	**0.0030**	**1**	**249289**	**71**	**61.88**	**0.0000**	1	201995	3	55.23	0.0000
2	53147	94	46.88	0.0040	2	252328	61	42.58	0.0140	2	176710	68	40.92	0.0030	2	269840	9	74.85	0.0000
3	157856	28	28.48	0.0060	3	260665	40	40.08	0.0170	3	196908	17	31.55	0.0080	3	270792	136	77.41	0.0000
4	126500	32	23.95	0.0180	**4**	**249289**	**158**	**52.16**	**0.0260**	4	108685	53	42.94	0.0060	4	77127	124	73.43	0.0005
5	138878	29	21.56	0.0410	5	176796	177	43.48	0.0430	5	176796	113	43.75	0.0110	5	166901	124	59.11	0.0010
6	72320	16	28.01	0.0450	6	196908	27	32.22	0.0930	6	57282	12	32.99	0.0190	6	248665	101	44.66	0.0055
7	158003	45	25.50	0.0470	**7**	**249281**	**7**	**28.15**	**0.1030**	7	122424	121	42.11	0.0110	7	154504	231	80.53	0.0075
8	63363	46	25.37	0.0510	8	48674	85	30.14	0.1440	8	228010	72	33.86	0.0160	8	235024	116	48.25	0.0075
9	106097	25	22.51	0.0630	9	254690	173	35.62	0.1750	9	217921	36	29.70	0.0200	9	267139	32	46.09	0.0090
10	96687	27	19.34	0.0680	10	119059	18	23.60	0.1630	10	168270	47	29.92	0.0240	10	**270857**	**43**	**60.28**	**0.0120**
11	113425	94	26.19	0.0800	11	217085	136	37.32	0.1940	11	249637	11	30.13	0.0280	11	270423	124	45.39	0.0130
12	125830	20	21.22	0.0890	12	257613	143	34.36	0.2050	12	249335	20	23.33	0.0260	12	263257	60	45.49	0.0140
13	131563	33	19.18	0.0930	13	263184	71	30.04	0.2170	13	205273	28	25.06	0.0300	13	263052	128	44.89	0.0150
14	127566	58	31.81	0.0970	14	89370	271	33.66	0.2080	14	252328	35	28.46	0.0340	14	230138	108	44.65	0.0155
15	154016	26	19.35	0.1000	15	166901	137	34.49	0.2360	15	239232	38	28.75	0.0320	15	114960	48	54.40	0.0160
16	353	4	18.24	0.1000	16	215899	109	31.60	0.2790	16	48674	22	21.42	0.0350	16	228001	183	65.23	0.0200
17	167525	24	20.69	0.1020	17	250376	89	25.23	0.2720	17	215899	63	34.65	0.0350	17	247960	102	43.54	0.0225
18	77208	33	18.62	0.1040	18	249335	41	22.28	0.2860	18	187593	27	26.07	0.0330	18	210277	24	43.41	0.0245
19	3654	226	41.93	0.1040	19	236937	9	23.92	0.3280	19	257613	75	30.93	0.0300	19	104043	20	32.25	0.0275
20	132632	36	19.09	0.1080	20	236608	258	25.33	0.3080	20	212342	56	39.03	0.0370	20	221155	92	42.01	0.0290
21	36846	33	18.43	0.1090	21	122424	13	19.27	0.3570	21	253502	31	26.04	0.0460	21	234567	19	33.32	0.0305
22	101877	33	18.09	0.1190	22	165441	83	24.78	0.4490	22	148334	43	31.37	0.0410	22	208310	74	50.87	0.0335
23	110841	11	23.76	0.1200	23	240761	178	25.93	0.4720	23	252514	30	24.05	0.0450	23	272736	22	40.29	0.0345
24	165085	37	20.14	0.1220	24	245534	105	26.91	0.4740	24	152072	42	31.83	0.0530	24	267007	23	39.38	0.0360
25	136760	170	42.35	0.1260	25	228818	90	21.15	0.5020	25	77127	14	27.21	0.0590	25	228028	31	33.10	0.0445
26	161365	45	21.13	0.1340	26	252689	81	25.95	0.4850	26	216380	25	21.98	0.0780	26	176796	23	34.65	0.0515
27	111066	7	22.00	0.1380	27	138118	159	26.91	0.5360	27	248789	21	18.56	0.0790	27	3336	138	48.98	0.0540
28	13200	37	19.50	0.1430	28	199907	103	26.05	0.5280	28	246689	56	34.83	0.0720	28	221350	87	37.71	0.0545
29	166650	29	17.15	0.1440	29	177393	84	22.87	0.4970	29	188956	52	28.49	0.0840	29	252328	138	48.28	0.0625
30	153052	24	18.81	0.1540	30	232785	111	24.20	0.5340	30	176869	11	23.05	0.0910	30	272612	108	37.03	0.0695
31	128660	61	24.91	0.1570	31	233170	95	23.25	0.5650	31	50202	69	25.54	0.0790	31	124915	78	44.07	0.0725
32	165441	76	28.40	0.1660	32	246077	27	19.59	0.5990	32	92649	48	25.89	0.0910	32	253162	18	28.16	0.0735
33	39519	93	21.89	0.1740	33	194883	33	18.23	0.5940	33	154504	26	19.96	0.1330	33	222763	27	33.88	0.0780
34	122165	45	19.50	0.1810	34	250201	67	19.50	0.6000	34	233897	22	16.38	0.1290	34	269204	107	35.43	0.0830
35	33766	37	18.11	0.1820	35	153869	272	23.00	0.6050	35	204145	36	21.02	0.1270	35	217085	106	35.58	0.0850
36	100208	7	17.68	0.1850	36	190691	43	14.99	0.6050	36	254690	69	23.20	0.1200	36	246174	53	36.45	0.0850
37	131679	90	22.30	0.1910	37	168270	74	21.55	0.6390	37	180548	51	23.56	0.1350	37	147524	48	41.87	0.0945
38	131180	50	19.68	0.1920	38	226904	12	16.42	0.6390	38	245500	10	22.76	0.1250	38	242268	116	33.77	0.0945
39	84778	137	28.20	0.1960	39	242560	18	14.67	0.6330	39	125687	28	18.45	0.1400	39	254134	62	31.04	0.1005
40	158844	7	15.79	0.1990	40	246182	38	17.17	0.6760	40	220221	62	27.03	0.1350	40	267511	93	33.58	0.1020
41	50202	213	40.14	0.2010	41	223484	36	20.27	0.6770	41	253367	26	19.38	0.1470	41	21857	114	32.68	0.1110
42	136727	28	15.59	0.2010	42	243728	100	22.86	0.6860	42	260665	32	21.28	0.1770	42	56871	43	42.16	0.1200
43	29440	21	16.31	0.2010	43	230510	107	21.90	0.6740	43	250295	6	24.85	0.1540	43	253022	106	32.94	0.1285
44	158569	36	16.71	0.2040	44	233706	72	19.20	0.6820	44	198439	52	21.67	0.2130	44	202622	71	40.77	0.1330
45	1570	46	18.59	0.2070	45	247162	76	20.45	0.6910	45	251070	52	21.45	0.2190	45	259241	12	30.90	0.1375
46	93998	73	27.81	0.2120	46	215791	31	15.43	0.6930	46	240761	119	23.46	0.2720	46	70378	28	28.30	0.1420
47	157325	22	15.67	0.2270	47	149624	48	16.51	0.6920	47	194883	8	20.92	0.2360	47	251160	62	29.20	0.1475
48	149640	27	14.90	0.2270	48	247677	146	20.74	0.7190	48	176818	18	13.47	0.2640	48	49956	96	30.89	0.1560
49	27	22	15.56	0.2320	49	28282	104	20.67	0.7280	49	215686	15	16.47	0.2820	49	249025	92	30.8[illegible]	0.1585
50	980340	45	18.11	0.2340	50	246417	94	19.42	0.7330	50	185124	74	18.56	0.3070	50	101508	146	42.90	0.1590

Source: SOEP, individual questionnaire, only monetary variables (own calculation)

3.2 Results with the Variability Method

Like in the Benford's Law approach the variability methods calculates a plausibility-value for each interviewer. The procedure is illustrated in figure 6. The value of T_i (as defined in equation 2), that is assigned to interviewer I_i, is compared to the corresponding distribution of the test statistic T, which is estimated by a resampling approach. The area under the density curve on the left side of the realization T_i defines the plausibility. If the plausibility is too small, the interviewer is considered to be a potential faker. The procedure corresponds to an one-sided statistical test. One could argue that interviewers who achieve a plausibility that is suspicious large, could be fakers as well. Following this argumentation one has to conduct a two-sided test. However, there is empirical evidence that this argument does not hold and that for the given task an one-sided statistical test is more appropriate. In table 7 several interviewer rankings (SOEP Sample AB, wave 1, SOEP Sample E, wave 1 + 2, SOEP Sample F, wave 1) are shown. The known fakers appear at the beginnings of the rankings. It is remarkable, that the interviewer 249289, who had faked questionnaires in two waves of Sample E and who was detected only in the third wave, is immediately debunked with the variability method in wave 1. Notice, that in all shown rankings the first rank is shared by two or more interviewers. Interviewers who achieve the same plausibility value are sorted in increasing order of their personal identification number.

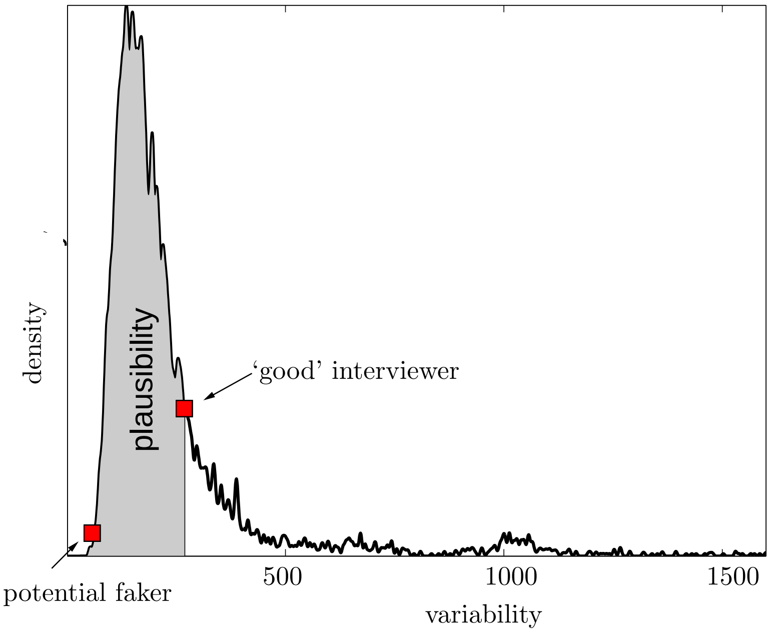

Figure 6: Example of the derivation of the plausibility for two realizations of the variability test statistic T: one interviewer achieves a very small plausibility and is therefore regarded as a potential faker, while the other achieves a good plausibility value. The depicted density is calculated on Sample AB, wave 1.

Table 7: *Interviewer-ranking with the variability method (faking interviewer bold)*

Sample AB				Sample E								Sample F			
wave 1				wave 1				wave 2				wave 1			
Rank	Int.no.	Q.no.	plausibility	Rank	Int.no.	Q.no.	plausibility	Rank	Int.no.	Q.no.	plausibility	Rank	Int.no.	Q.no.	plausibility
1	16306	25	0.00254	1	50202	25	0.00000	**1**	**249289**	22	0.00000	1	53104	55	0.00000
1	33111	222	0.00254	1	138118	27	0.00000	1	246077	2	0.00000	1	240290	74	0.00000
1	33766	40	0.00254	1	166901	29	0.00000	3	167240	20	0.00016	1	253723	17	0.00000
1	103012	89	0.00254	**1**	**249289**	25	0.00000	4	230855	2	0.00026	1	167240	29	0.00000
1	157856	18	0.00254	1	254690	29	0.00000	5	138118	24	0.00058	1	237515	28	0.00000
1	165441	29	0.00254	1	260665	12	0.00000	6	215686	6	0.00138	1	252514	14	0.00000
7	152870	22	0.00252	7	250201	10	0.00024	7	154890	10	0.00302	1	253260	11	0.00000
8	**139378**	32	0.00254	**8**	**249281**	2	0.00044	8	250201	10	0.00418	8	264016	20	0.00012
9	64343	35	0.00258	9	165441	19	0.00054	9	249637	3	0.00444	**9**	**270857**	11	0.00016
10	**128279**	35	0.00259	9	167240	24	0.00054	10	199907	35	0.00536	10	242691	2	0.00018
11	89370	119	0.00266	11	120820	25	0.00064	11	251275	13	0.00856	11	270750	18	0.00034
12	36145	22	0.00281	12	240290	71	0.00114	12	233706	7	0.00914	12	212288	61	0.00044
13	149624	64	0.00317	13	205273	22	0.00164	13	166901	30	0.01264	13	112348	20	0.00056
14	**43800**	38	0.00323	13	253502	18	0.00164	14	167371	5	0.01382	13	201995	2	0.00056
15	167916	13	0.00338	15	199907	27	0.00174	15	233862	17	0.01476	15	251160	13	0.00058
16	118320	6	0.00344	16	252328	15	0.00224	16	244333	12	0.02394	16	263257	15	0.00084
17	29440	13	0.00345	**17**	**236837**	32	0.00324	17	254690	29	0.03044	17	254134	15	0.00088
18	166901	14	0.00363	18	89370	49	0.00384	18	252689	9	0.03376	18	259756	19	0.00092
19	169161	11	0.00373	19	217086	2	0.00634	19	176710	35	0.03588	19	267872	17	0.00094
20	53104	66	0.00382	20	251275	15	0.00714	20	204145	14	0.04594	20	272043	18	0.00168
21	164704	2	0.00399	21	204145	14	0.00874	21	112348	10	0.05626	21	249548	54	0.00204
22	167460	8	0.00427	22	233862	27	0.00914	22	194883	4	0.06354	22	188956	38	0.00212
23	105473	33	0.00443	23	250376	12	0.01074	23	252328	12	0.06984	23	251089	16	0.00286
24	51187	6	0.00445	24	177393	15	0.01174	24	255149	5	0.07934	24	250910	10	0.00314
25	158747	60	0.00474	25	217921	9	0.01344	25	65030	19	0.07944	25	263575	20	0.00364
26	130206	24	0.00477	26	39160	13	0.01674	26	185302	2	0.08004	26	260045	15	0.00384
27	165093	11	0.00506	27	226904	3	0.02554	27	177393	15	0.08446	27	263427	4	0.00402
28	103730	30	0.00549	28	190691	8	0.02724	28	108685	14	0.09642	28	259683	13	0.00428
29	980340	10	0.00579	29	246689	19	0.02774	29	253502	14	0.10844	29	254649	6	0.00442
30	39160	29	0.00599	30	239330	9	0.03714	30	205273	12	0.11182	30	234630	29	0.00448
31	104043	49	0.00640	31	246379	17	0.04224	31	232785	12	0.11398	31	250406	19	0.00484
32	72320	3	0.00658	32	235083	15	0.04374	32	48674	7	0.12314	32	253430	23	0.00518
33	159891	15	0.00733	33	215899	13	0.04644	33	233846	5	0.12916	33	239356	3	0.00532
34	7773	88	0.00741	34	227870	6	0.05144	34	230413	26	0.13236	34	264660	31	0.00552
35	160121	21	0.00801	35	168270	8	0.05174	35	160032	4	0.13934	35	262030	16	0.00574
36	117889	8	0.00834	36	195898	23	0.05864	36	227870	6	0.14106	36	260762	2	0.00584
37	48135	36	0.00838	37	255009	4	0.05944	37	239232	9	0.15502	37	246174	10	0.00588
38	168157	7	0.00859	38	257249	23	0.05974	38	215791	5	0.16474	38	224596	11	0.00604
39	4103	3	0.00874	39	215791	7	0.06454	39	252514	7	0.18984	39	219789	9	0.00646
40	59749	7	0.00959	40	259845	21	0.06684	40	119806	14	0.20304	40	253022	27	0.00654
41	113719	42	0.01023	41	170070	2	0.06944	41	250473	2	0.21638	41	191779	20	0.00694
42	64254	11	0.01037	42	119806	16	0.07194	42	212342	14	0.21724	42	133825	24	0.00712
43	2746	2	0.01053	43	253316	5	0.07294	43	201340	17	0.22452	43	270660	14	0.00848
44	89770	6	0.01055	44	233170	13	0.07574	44	235083	14	0.24428	44	253375	22	0.00934
45	136760	55	0.01065	45	263184	7	0.07624	45	250376	17	0.27914	45	252816	10	0.01044
46	38423	24	0.01141	46	259849	1	0.08284	46	253294	9	0.28794	46	270792	47	0.01064
47	131890	30	0.01154	47	226076	17	0.09864	47	196908	3	0.28902	47	247294	106	0.01086
48	161110	49	0.01184	48	262060	1	0.10124	48	120820	24	0.29026	47	270954	25	0.01086
49	19852	154	0.01235	49	252891	27	0.10694	49	259845	18	0.29076	49	253162	6	0.01088
50	157465	13	0.01283	50	154504	3	0.11434	50	249335	5	0.29168	50	264717	23	0.01098

Int.no.: number of interviewers, Q.no.: number of questionnaires

Source: SOEP, individual questionnaire (own calculation)

4 Impact of suspicious interviewers on results

In this section we present some empirical results using suspicious interviewer clusters identified by Benford's Law and the variability method as well as already identified evident fabricated interviews and assumed non-faked data. We will look first at some descriptive statistics like proportions, means and variances. We analyze only samples A, B and E because the number of detected fakes in subsample F is too small (N= 8). While we are able to give valid values for the maximal possible bias in the case of means and proportions the given values for the empirical bias are only estimates under the assumption that the distribution of the unknown true data follows the assumed non-faked or non-suspicious data.

4.1 Bias

Estimation of the possible bias due to interviewer cheating. The possible bias due to falsifications is formally similar to the possible bias due to imputation of values in the case of missing data. We can interpret falsifications as a special kind of imputation that depends on an interviewer's assumptions about an unknown respondent's characteristics and opinions (cf. Schnell 1991).

In this section we show some simple equations for calculating the possible bias due to interviewer cheating[1] (following Kalton 1983, p.6-10 and extended for our problem). For simplicity we will consider a simple random sample of size n drawn from a population of size N, and we will first concentrate on a single variable Y. Let N_{nf} be the number of non-faked interviews and N_f be the number of fabricated interviews in the population, with $N = N_{nf} + N_f$. The corresponding sample quantities are n_{nf} and n_f, with $n = n_{nf} + n_f$. The population total are given by $Y = Y_{nf} + Y_t$, and the population mean is given by $M = N_{nf}/N \cdot M_{nf} + N_f/N \cdot M_t$, where Y_{nf} and M_{nf} are the total and mean for non-faked data and Y_t and M_t are the same quantities for the nonrespondents. The corresponding sample quantities are $y = y_{nf} + y_t$ and $m = n_{nf}/n \cdot m_{nf} + n_f/n \cdot m_t$. Cheating interviewers try to imputate the missing values of the nonrespondents with faked data in the sample. The sample quantities y_f and m_f are the total and the mean for the faked data. If the faked data are not detected the sample mean $m_w = n_{nf}/n \cdot m_{nf} + n_f/n \cdot m_f$ contains not the true but the faked quantities for the nonrespondents. This sample mean m_w is used to estimate the population mean M. Its bias is given by $B(m_w) = E(m_w) - M$. The expectation of m_w is

$$E(m_w) = E\left(\frac{N_{nf}}{N}E(M_{nf}) + \frac{N_f}{N}E(M_f)\right) = \frac{N_{nf}}{N}M_{nf} + \frac{N_f}{N}M_f$$

Hence the bias of the mean m_w is given by

$$B(m_w) = M_w - M = \frac{N_{nf}}{N}M_{nf} + \frac{N_f}{N}M_f - \frac{N_{nf}}{N}M_{nf} - \frac{N_f}{N}M_t = \frac{N_f}{N}(M_f - M_t) \quad (4)$$

[1] Schnell (1991) gives also equations for the possible bias in samples with fabricated data. We don't use them because he makes the implicit assumption that the data for the faked cases are the same as the data for the non-faked part of the sample. He doesn't explicit distinguish between the unknown true data and the non-faked data.

Equation 4 shows that m_w is approximately unbiased for M if either the proportion of the fakes N_f/N is small or the mean in the fabricated data is close to that for the unknown true data M_t. Unfortunately we have no direct empirical evidence on the magnitude of $(M_f - M_t)$. If we assume that the unknown true values y_t have the same distribution like the known non-faked values y_{nf}, we will get a rough estimate for the empirical bias $B(m_w)|(M_t = M_{nf}) = N_f/N(M_f - M_{nf})$.

In the case of proportions the bias is given by

$$B(p_w) = P_w - P = \frac{N_f}{N}(P_f - P_t) \tag{5}$$

Equation 5 shows that the bias for the proportion can't be greater than the proportion of the falsified values in the sample. Hence, if there are 3% fakes in the sample, the maximum bias can be no more than 3%. Again, if we assume equal distribution for non-faked and true data, we will get an estimate of the empirical bias with $B(p_w)|(P_t = P_{nf}) = N_f/N(P_f - P_{nf})$.

Finally, we consider the effect of cheating on the estimation of variances and covariances. The expectation of the respondent sample variance s_w^2 is $E(s_w^2) = E(S_w^2) = S_w^2$ where

$$S_w^2 = \frac{N_{nf}}{N}S_{nf}^2 + \frac{N_f}{N}S_f^2 + \frac{N_{nf}}{N}\frac{N_f}{N}(M_{nf} - M_f)^2$$

The bias of s_w^2 as an estimator for S^2 is thus $B(s_w^2) = S_w^2 - S^2$ where

$$S^2 = \frac{N_{nf}}{N}S_{nf}^2 + \frac{N_f}{N}S_t^2 + \frac{N_{nf}}{N}\frac{N_f}{N}(M_{nf} - M_t)^2$$

Hence the bias is

$$B(s_w^2) = \frac{N_f}{N}(S_f^2 - S_t^2) + \frac{N_{nf}}{N}\frac{N_f}{N}[(M_{nf} - M_f)^2 - (M_{nf} - M_t)^2] \tag{6}$$

The first term of this bias is comparable to the bias for a mean and the proportion in 4 and 5. The second term reflects the effect of differences in the non-faked and faked mean as well as the true mean on the estimator. Under the assumptions of equal distribution for non-faked and true data so that $S_{nf}^2 = S_t^2$ and $M_{nf} = M_t$, we will get an estimate for the empirical bias with $B(s_w^2)|(S_{nf}^2 = S_t^2, M_{nf} = M_t) = N_f/N(S_f^2 - S_{nf}^2) + N_{nf}/NN_f/N(M_{nf} - M_f)^2$.

For the covariance another variable, x, needs to be introduced. We assume that respondents provide both x and y values. If the expectation of the sample covariance s_{xy_w} is $E(S_{xy_w}) = S_{xy_w}$ where

$$S_{xy_w} = \frac{N_{nf}}{N}S_{xy_{nf}} + \frac{N_f}{N}S_{xy_f} + \frac{N_{nf}}{N}\frac{N_f}{N}(M_{x_{nf}} - M_{x_f})(M_{y_{nf}} - M_{y_f})$$

The bias of s_{xy_w} as an estimator of S_{xy} is $B(s_{xy_w}) = S_{xy_w} - S_{xy}$ where

$$S_{xy} = \frac{N_{nf}}{N}S_{xy_{nf}} + \frac{N_f}{N}S_{xy_t} + \frac{N_{nf}}{N}\frac{N_f}{N}(M_{x_{nf}} - M_{x_t})(M_{y_{nf}} - M_{y_t})$$

and the bias is

$$B(s_{xy_w}) = \frac{N_{nf}}{N}(S_{xy_f} - S_{xy_t}) + \frac{N_{nf}}{N}\frac{N_f}{N}[(M_{x_{nf}} - M_{x_f})(M_{y_{nf}} - M_{y_f}) - (M_{x_{nf}} - M_{x_t})(M_{y_{nf}} - M_{y_t})] \tag{7}$$

Empirical results. We build for Sample A and E three subsamples of faked or suspicious data "identified" by Benford or the variability method to explore the impact on empirical result. The suspicious interviewer clusters occur on the top of the interviewer-rankings in the former section. Nevertheless, it is difficult to make a clear discrimination between suspicious and non-suspicious clusters. The usage of a statistical criterion like a 5% level for the estimated plausibility will result in some cases obviously in too many suspicious interviewer clusters. A more pragmatic decision is to declare the first ten interviewers with the lowest plausibility as suspicious. Table 8 shows the selected suspiciously interviewers with their rank. The percentage of suspicious individual interviews is shown in table 9.

TABLE 8: *Selected suspiciously interviewers*

	evident	Benford	Rank	variability	Rank
Sample A, w1	128279	53147	2	16306	1
	43800	126500	4	33111	1
	139378	138878	5	33766	1
		72320	6	103012	1
		157856	3	157856	1
		158003	7	165441	1
		63363	8	152870	7
		106097	9	64343	9
Sample E, w1	249289	252328	2	50202	1
	236837	260665	3	260665	1
	249281	176796	5	254690	1
		196908	6	166901	1
		48674	8	138118	1
		254690	9	250201	7
		119059	10	165441	9
				167240	9
				120820	11

Next the estimates based on each of these suspicious sub-sample are compared with estimates based on non-faked respectively non-suspicious data.

Proportions. In the previous section we demonstrated that the possible bias can not be greater than the proportion of falsified values in the sample. The next three tables show proportions and frequencies of some selected variables.

Table 10 shows the breakdown of gender responses in fabricated and real samples. We can detect in all cases only a marginal empirical bias. It can be assumed that it is rather easy for cheating interviewers to reproduce responses like respondent's

Table 9: *Percentage of faked or suspiciously interviewers in the sample A/B and E*

	evident fakes		Benford		Variability	
	N	%	N	%	N	%
Sample A, w1						
non-fake	9115	99.40	9009	98.84	8893	97.56
fake	59	0.60	106	1.16	222	2.44
total	9174	100.00	9115	100.00	9115	100.00
Sample B, w1						
non-fake	3175	98.60	3161	99.56	2917	91.87
fake	45	1.40	14	0.44	258	8.13
total	3220	100.00	3175	100.00	3175	100.00
Sample E, w1						
non-fake	1910	97.60	1766	92.50	1710	89.50
fake	47	2.40	144	7.50	200	10.50
total	1957	100.00	1910	100.00	1910	100.00

gender because the distribution is known. Hence we will take a look at other variables with more categories. It might be a somewhat more complicated to reproduce the employment status of the SOEP respondents.

Table 10: *Proportion of respondent's gender in fabricated and non-faked as well as in suspicious and non suspicious data*

	evident		emp.	*Benford*		emp.	*Variability*		emp.
	non-fake	fake	Bias	non-susp.	susp.	Bias	non-susp.	susp	Bias
Sample A, wave 1									
male	47.69	45.76	-0.012	47.66	50.00	0.027	47.70	47.30	-0.010
female	52.31	54.24	0.012	52.34	50.00	-0.027	52.30	52.70	0.010
total	100.00	100.00		100.00	100.00		100.00	100.00	
Sample B, wave 1									
male	52.98	62.22	0.129	52.96	42.86	-0.045	53.14	51.16	-0.161
female	47.02	37.78	-0.129	47.04	57.14	0.045	46.86	48.84	0.161
total	100.00	100.00		100.00	100.00		100.00	100.00	
Sample E, wave 1									
male	48.80	55.32	0.157	48.58	51.39	0.211	48.48	51.50	0.316
female	51.20	44.68	-0.157	51.42	48.61	-0.211	51.52	48.50	-0.316
total	100	100		100	100.00		100	100.00	

Table 11 shows the distribution of respondent's employment status in samples A and B for fabricated and non-faked as well as suspicious data. This variable has seven categories. The highest frequency occurs for the category "full-time employment" with

46.7%, followed by "not employed" with 37.6%. Regular part-time employment responses are only 5.4%, followed by vocational training and unemployed, both with 3.6%. Surprisingly, the distribution for the faked sample and the suspicious samples are quite similar to the assumed non-faked data[2]. The ranking order of the categories corresponds in both data sets and there are only small deviations in the frequency values, especially for "full-time employment" and "not employed". Therefore we can expect that the cheating interviewers have an idea of the distribution of the employment status in the entire population and are able to reproduce the frequencies of this variable.

Table 11: *Distribution of employment status in Sample A + B, 1984*

	evident non-fake	fake	Benford non-susp.	susp.	Variability non-susp.	susp.
full-time employment	46.75	54.81	46.77	44.17	46.48	53.33
reg. Part-time employment	5.40	8.65	5.42	3.33	5.45	4.17
vocational training	3.59	3.85	3.55	6.67	3.62	2.71
marginal part-time employment	2.78	0.96	2.77	4.17	2.80	2.29
unemployed	3.59	4.81	3.61	0.83	3.63	2.50
military, civil service	0.29	0.00	0.29	0.00	0.30	0.00
not employed	37.62	26.92	37.58	40.83	37.72	35.00
total	100.00	100.00	100.00	100.00	100.00	100.00
N	12245	104	12125	120	11765	480

Source: SOEP, Sample A and B, non-faked and faked data, individual questionnaire

Table 12 shows the frequency of another item in samples A and B, "the importance of goals in politics".[3] The respondent has to choose between four goals and has to rank these in terms of personal importance. In the non-faked data set we can recognize that the distribution of "peace and quiet" has its highest frequency for the most important goal, "inflation" for the second most important goal, "citizen influence" has equal frequencies for the third and fourth most important goal, and "freedom of speech" has its highest frequency for the fourth most important goal. In the evident fabricated data we find a rather different distribution, by far the highest frequency for the first goal occurs for "inflation", "peace and quiet" has its highest frequency for the second goal, "freedom of speech" for the third, and "citizen influence" for the fourth.

4.2 Means and Variances

Table 13 shows some means and variances in fabricated and real data. We have calculated the means for the items 'importance for satisfaction' (4-point scale). The empirical bias in all cases is rather low and negligible; only half of the differences in

[2] A chi-square test shows that the difference is neither on a 1% or 5% nor 10% level significant (chi-square value is 8.674).

[3] The question is: "Even in politics you can't have everything at once. Below are various goals which politics can aim for; if you had to choose between these goals: which seems the most important to you? Which is the second most important? Which is the third most important? And, which is the fourth?"

Table 12: *Frequency of importance of goals in politics - Sample A + B*

	peace and quiet		*citizen influence*		*inflation*		*freedom of speech*	
	non-fake	fake	non-fake	fake	non-fake	fake	non-fake	fake
evident fakes								
1st. job	47.2	**23.1**	16.9	22.3	22.8	**50.0**	18.3	**4.9**
2nd. job	22.6	**41.3**	24.2	16.5	31.1	**26.0**	22.1	**16.5**
3rd. job	15.7	**7.7**	29.3	29.1	27.4	**22.1**	25.7	**40.8**
4th. Job	14.6	**27.9**	29.7	32.0	18.8	**1,9**	34.0	**37.9**
	100.0	**100.0**	100.0	100.0	100.0	**100.0**	100.0	**100.0**
N	11973	104	11865	103	11951	104	11928	103
	non-susp.	susp.	non-susp.	susp.	non-susp.	susp.	non-susp.	susp.
Benford								
1st. job	47.2	41.0	16.8	19.6	22.8	24.3	18.3	18.9
2nd. job	22.5	29.1	24.2	18.8	31.1	28.7	22.0	27.9
3rd. job	15.7	16.2	29.3	25.0	27.3	30.4	25.7	23.4
4th. Job	14.6	13.7	29.6	36.6	18.8	16.5	34.0	29.7
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	11856	117	11753	112	11836	115	11817	111
	non-susp.	susp.	non-susp.	susp.	non-susp.	susp.	non-susp.	susp.
Variability								
1st. job	47.4	40.9	17.3	**6.5**	22.0	41.8	18.5	11.9
2nd. job	22.1	34.7	24.8	**8.0**	30.9	35.5	22.1	21.8
3rd. job	15.8	13.4	29.4	**26.2**	27.9	15.7	24.9	44.6
4th. Job	14.7	11.1	28.5	**59.3**	19.2	7.1	34.5	21.8
	100.0	100.0	100.0	**100.0**	100.0	100.0	100.0	100.0
N	11494	479	11388	477	11472	479	11450	478

Source: SOEP, Sample A and B, non-faked and faked data, individual questionnaire

means between faked and non-faked respectively suspicious and non-suspicious data are significant.

The "fit" of the fabricated data are rather good, the absolute deviation between non-fake and faked mean is only 0.19 on average (Benford 0.17; Variability method 0.13). An exception is the interviewer's assessment of the importance of "work" for respondent's satisfaction. Here in the evident fabricated and the suspicious data the mean is 25% higher (variability method 15%) than in the non-fake respectively non-suspicious data.

Table 13: *Means and variances in fabricated and real data (Sample E)*

importance	evident fakes						Benford						Variability method					
for satisf.	means			F-Test		emp.	means			F-Test		emp.	means			F-Test		emp.
(4 point scale)	non-fake	fake	total	prob.		bias	non-susp.	susp.	total	prob.		bias	non-susp.	susp.	total	prob.		bias
work	1.92	2.41	1.93	0.000	***	0.01	1.88	2.40	1.92	0.000	***	0.04	1.89	2.18	1.92	0.000	***	0.03
family	1.28	1.06	1.27	0.004	***	-0.01	1.28	1.28	1.28	0.979		0.00	1.27	1.32	1.28	0.207		0.01
friends	1.72	1.34	1.71	0.000	***	-0.01	1.71	1.76	1.72	0.400		0.01	1.73	1.61	1.72	0.009	***	-0.01
income	1.61	1.72	1.61	0.193		0.00	1.61	1.67	1.61	0.199		0.00	1.61	1.60	1.61	0.852		0.00
lodging	1.57	1.62	1.57	0.533		0.00	1.57	1.53	1.57	0.437		0.00	1.56	1.59	1.57	0.554		0.01
politics	2.83	2.91	2.83	0.493		0.00	2.83	2.88	2.83	0.457		0.00	2.83	2.85	2.83	0.739		0.00
career	2.26	2.51	2.26	0.090	*	0.00	2.23	2.56	2.26	0.000	***	0.03	2.21	2.61	2.26	0.000	***	0.05
leisure	1.72	1.53	1.72	0.043	**	0.00	1.70	1.99	1.72	0.000	***	0.02	1.72	1.68	1.72	0.383		0.00
health	1.15	1.00	1.14	0.007	***	-0.01	1.15	1.15	1.15	0.818		0.00	1.15	1.13	1.15	0.392		0.00
environment	1.71	1.83	1.71	0.193		0.00	1.69	1.96	1.71	0.000	***	0.02	1.70	1.75	1.71	0.350		0.01
religion	2.68	2.43	2.67	0.079	*	-0.01	2.66	2.85	2.68	0.029	**	0.02	2.71	2.42	2.68	0.000	***	-0.03
neighborhood	1.88	1.85	1.88	0.739		0.00	1.87	2.01	1.88	0.005	***	0.01	1.86	2.00	1.88	0.001	***	0.02
mobility	1.78	1.98	1.79	0.049	**	0.01	1.76	2.00	1.78	0.000	***	0.02	1.76	1.96	1.78	0.000	***	0.02

SOEP, Sample Sample E, wave 1, signif.: * 10% level, ** 5% level, *** 1%level (own calculation)

importance	evident fakes				Benford				Variability method			
for satisf.	variances			emp.	variances			emp.	variances			emp.
(4 point scale)	non-fake	fake	total	bias	non-fake	fake	total	bias	non-fake	fake	total	bias
work	0.817	0.992	0.826	0.009	0.672	1.280	0.817	0.145	0.788	1.011	0.817	0.029
family	0.263	0.061	0.259	-0.004	0.265	0.231	0.263	-0.002	0.265	0.240	0.263	-0.002
friends	0.383	0.316	0.385	0.002	0.390	0.299	0.383	-0.007	0.390	0.310	0.383	-0.007
income	0.347	0.291	0.346	-0.001	0.351	0.307	0.347	-0.004	0.352	0.311	0.347	-0.005
lodging	0.305	0.372	0.306	0.001	0.308	0.265	0.305	-0.003	0.311	0.254	0.305	-0.006
politics	0.715	0.558	0.711	-0.004	0.716	0.702	0.715	-0.001	0.709	0.769	0.715	0.006
career	0.986	1.074	0.989	0.003	0.970	1.093	0.986	0.016	0.939	1.240	0.986	0.047
leisure	0.394	0.428	0.396	0.002	0.374	0.574	0.394	0.020	0.397	0.369	0.394	-0.003
health	0.138	0.000	0.135	-0.003	0.135	0.173	0.138	0.003	0.138	0.141	0.138	0.000
environment	0.393	0.362	0.393	0.000	0.386	0.421	0.393	0.007	0.387	0.452	0.393	0.006
religion	0.937	0.858	0.936	-0.001	0.955	0.680	0.937	-0.018	0.942	0.827	0.937	-0.005
neighborhood	0.320	0.130	0.315	-0.005	0.317	0.345	0.320	0.003	0.320	0.305	0.320	0.000
mobility	0.463	0.543	0.465	0.002	0.459	0.454	0.463	0.004	0.453	0.514	0.463	0.010

SOEP, 1998, Sample E (own calculation) signif.: * 10% level, ** 5% level, *** 1%level

4.3 Correlations

In this section we will examine the influence of fabricated data on bivariate statistics like covariances and correlations. Table 14 shows the correlation between net and gross income as well as between gross income and duration of training (years) in sample A and E. The relationship between gross and net income is trivial and obvious and apparent in both non-faked and fabricated as well as suspicious and non-suspicious data. However the connection with "duration of training" (generated from the variable for schooling and training in years) is more complicated and more adjustments are required. On the basis of human capital theory we expect a positive correlation and find a significant positive value of 0.367 (sample A) and 0.342 (sample E) in the assumed non-faked data that doesn't contain the evident fabricated data.

For the evident fabricated data we get partly inconsistent results, in sample A the correlation is 0.470 and not significant higher than in the non-faked data and in sample E only a small negative insignificant correlation occurs. Although the amount of fakes in sample E is under 5% and very small, the impact of the fakes in the overall sample on the correlation is serious, biasing the total positive correlation downward to a value of 0.271.

After deleting the evident fakes in sample A and E we found in the non-suspicious data selected by the Benford and variability method equal correlations as in the evident non-faked data. But in the suspicious data set of sample A we can recognize a clearly lower correlation with training based on Benford and a higher correlation for the suspicious data of the variability method.

TABLE 14: *Correlation in fabricated and suspicious as well as non-suspicious data*

correlation between gross income and . . .	Sample A, w1			Sample E, w1		
	net income	training	N	net income	training	N
evident						
non-fake	0.943***	0.367***	4369	0.948***	0.342***	699
fake	0.989***	0.470***	32	0.924***	**-0.004**	27
total	0.943***	0.367***	4401	0.948***	0.271***	726
Benford						
non-suspicious	0.943***	0.369***	4319	0.948***	0.342***	589
suspicious (without evident)	0.965***	**0.140**	50	0.951***	0.324*	31
Variability						
non-suspicious	0.942***	0.364***	4266	0.947***	0.329***	558
suspicious (without evident)	0.974***	0.466***	103	0.969***	0.408***	62

Source: SOEP, Sample A and E, signif.: * 10% level, ** 5% level, *** 1%level

4.4 Linear regressions

In a further step we examine the impact of fakes on multivariate statistics like linear regressions. One of the most important regressions in a socio-economical context is the regression of log gross income. In our equation we use "age" (in years), "age squared", "gender", "duration of training" and "working hours per week" as right-hand variables. Tables 15 and 16 shows the estimated parameters for sample A (1984) and tables 17 and 18 for sample E (1998). In the assumed non-faked and non-suspicious samples all coefficients have the expected signs and they are significant, the log gross income increases with duration of training, working hours and the age of respondents (proxy for vocational experience), and male respondents have higher incomes than females. The coefficients are reasonable and the overall fit of these models is measured in both samples with adjusted $R^2 = 0.487$ and $R^2 = 0.542$. In the evident fabricated data sets (0.6% of the subsample A and 4.7% of the subsample E) we find inconsistent results. While in sample A the estimated parameters are rather close to which in the assumed non-faked data set (with exception of the overestimation of training), we find some differences in sample E. In sample E the coefficient of duration of training is negative and implausible, the coefficient for working hours is only a third and the coefficient for gender is more than double the coefficient in the non-faked data set.

If we leave the evident fabricated data in sample E we will get biased estimates. In the overall sample E concerning to the non-faked data the sign of the estimated parameters don't change but the covariates of "age" and "gender" are overestimated and "duration of training" and "working time" are underestimated. The overall fit is lower than in the non-faked data set, the value for adj. R^2 declines to 0.378.

The tables 16 and 18 show that there is also a minor effect of the suspicious interviewer cluster that are found with Benford and the variability method. The Gender effect in the suspicious data of Benford is more than triple than in the non-suspicious data and the training effect in the suspicious data of the variability method is more than twice than the non-suspicious data. Nevertheless the last column shows that these differences have only minor effects on the overall estimates.

TABLE 15: *Linear regression on gross log-income, Sample A, wave 1 - with/without evident fakes*

Sample A, w1 Regression on log-gross-income	non-fake coeff.	t- value	fake coeff.	t- value	total coeff.	t- value
const	3.417***	40.27	3.565***	4.33	3.421***	40.57
age	0.119***	29.22	0.089***	2.19	0.119***	29.33
age squared	-0.001***	-25.81	-0.001*	-1.75	-0.001***	25.90
sex (1 - men)	0.523***	29.84	0.457***	2.36	0.523***	29.95
duration of training (years)	0.089***	24.94	**0.137****	2.02	0.089***	25.02
working time (week)	0.014***	24.09	0.013*	1.98	0.014***	24.44
adj. R^2	0.487		0.578		0.488	
N	4353		29		4383	

Source: SOEP, Sample A, signif.: * 10% level, ** 5% level, *** 1% level
(own calculation)

5 Discussion

This paper deals with the identification and impact of fabricated interviews in the German Socio-Economic Panel (SOEP).

The data basis of this paper are the raw data of the German Socio-Economic Panel

Table 16: *Linear regression on gross log-income, Sample A, wave 1 - with/without suspicious interviewers detected by Benford and the variability method (without evident fakes)*

Sample A, w1 Regression on log-gross-income	Benford without evident				Variability without evident				total	
	non-suspicious		suspicious		non-suspicious		suspicious			
	coeff.	t-value	coeff.	t-value	coeff.	t-value	coeff.	t-value	coeff.	t-value
const	3.415***	39.981	3.532***	4.58	3.398***	39.63	4.501***	8.02	3.417***	40.27
age	0.118***	28.93	0.160***	4.48	0.120***	29.08	0.066**	2.45	0.119***	29.22
age squared	-0.001***	-25.55	-0.002***	-4.09	-0.001***	-25.7	-0.001*	-2.07	-0.001***	-25.81
sex (1 - men)	0.523***	29.65	0.456***	3.04	0.514***	29.01	**0.898*** **	7.89	0.523***	29.84
duration of training (years)	0.090***	24.95	0.029	0.984	0.089***	24.66	0.088***	3.90	0.089***	24.94
working time (week)	0.014***	23.96	0.014*	1.98	0.014***	24.11	0.007*	1.95	0.014***	24.09
adj. R^2	0.487		0.486		0.487		0.541		0.487	
N	4303		49		4250		102		4353	

Source: SOEP, Sample A, signif.: * 10% level, ** 5% level, *** 1% level
(own calculation)

Table 17: *Linear regression on gross log-income, Sample E, wave 1 - with/without evident fakes*

Sample E, w1 Regression on log-gross-income	non-fake		fake		total	
	coeff.	t-value	coeff.	t-value	coeff.	t-value
const	3.448***	13.33	5.539***	2.69	4.916***	18.2
age	0.111***	10.36	0.151	1.49	0.125***	10.2
age squared	-0.001***	-8.44	-0.002	-1.34	-0.001***	-8.74
sex (1 - men)	0.170***	3.86	**0.477*** **	2.87	0.306***	6.37
duration of training (years)	0.074***	9.39	**-0.029**	-0.67	0.014*	1.82
working time (week)	0.042***	15.34	0.015	0.96	0.021***	8.17
adj. R^2	0.542		0.296		0.378	
N	520		26		546	

Source: SOEP, Sample E, signif.: * 10% level, ** 5% level, *** 1% level
(own calculation)

(SOEP). A total of 90 faked household interviews and 184 faked individual interviews were detected by conventional verification methods, like reinterviewing, almost all of them in the first wave of a subsample. The share of fabricated data is low in all samples (far less than 1%) and the maximum is 2.4% in sample E. In subsamples C and D no fakes occurred. One should note that except for the fakes in sample E, faked data were never disseminated within the widely-used SOEP. The fakes were detected before the data were released. But those fakes are in the original data files which were provided by the fieldwork organization - kept at DIW Berlin - and they are a rich source for methodological research. Only one interviewer was able to fabricate interviews in the first two waves in sample E, thus they were detected after wave three. In other cases the faked data of wave 2 were not delivered by the fieldwork organization to DIW Berlin.

We applied two new approaches for discovering frauds which does not need two waves of data but which can be applied on cross-sections. First we found a recent practice becoming common among accountants, the Benford distribution of numbers, for fraud detection and assign this procedure to survey data. Second we use a new method called variability method that exploits the empirical observation that fakers

Table 18: *Linear regression on gross log-income, Sample E, wave 1 - with/without suspicious interviewers detected by Benford and the variability method (without evident fakes)*

Sample E, w1 Regression on log-gross-income	Benford without evident				Variability without evident				total	
	non-suspicious		suspicious		non-suspicious		suspicious			
	coeff.	t-value	coeff.	t-value	coeff.	t-value	coeff.	t-value	coeff.	t-value
const	3.446***	13.04	3.314**	2.27	3.540***	13.11	2.088**	2.26	3.448***	13.33
age	0.109***	10.03	0.167**	2.76	0.112***	10.17	0.091**	2.06	0.111***	10.36
age squared	-0.001***	-8.15	-0.002**	-2.41	-0.001***	-8.35	-0.001	-1.46	-0.001***	-8.44
sex (1 - men)	0.160***	3.55	**0.531****	2.19	0.173***	3.76	0.177	1.22	0.170***	3.86
duration of training (years)	0.074***	9.22	0.07*	1.63	0.068***	8.3	**0.160*****	4.59	0.074***	9.39
working time (week)	0.042***	15.16	0.029*	1.64	0.041***	14.43	0.061***	5.84	0.042***	15.34
adj. R^2	0.544		0.470		0.541		0.597		0.542	
N	492		27		468		51		520	

Source: SOEP, Sample E, signif.: * 10% level, ** 5% level, *** 1% level
(own calculation)

most often produces answers with less variability than could be expected from the whole survey.

In both procedures we derive test statistics for each interviewer cluster. The distributions of these test statistics are estimated using resampling approaches on the whole survey. From these distributions we can derive probabilities of the observed values. We sort the interviewers with respect to the plausibility they achieved and obtain an interviewer ranking. The interviewers with lowest plausibility are at the top of the ranking. They are considered to be potential fakers. We could show that we can identify with Benford as well as with the variability method the most clusters of interviews which were fabricated out of clusters which we know that they are faked.

To explore the impact of faked and suspicious interviews we build subsamples of non-faked and faked interviews that are already detected by the fieldwork organization as well as non-suspicious and suspicious interviews that are detected by Benford and the variability method. The share of evident faked or suspicious data is low in all samples and the maximum is 2.4% (evident fakes), 7.5% (Benford) and 10.6% (variability method) in sample E.

We show that the impact of interviewer cheating on proportions can not be greater than the proportion of the fakes in the sample. Under the assumption that the distribution of the unknown true data follows the known non-faked data we give estimates of the empirical bias. Overall we could observe that the estimated bias for proportions is very small and negligible in the SOEP, not only because the share of fakes is low, but because the "quality" of fakes is high. Interviewers who cheat often have an idea of the distribution of a particular variable such as "employment status" and can successfully reproduce the frequencies of this variable in the data they deliver to the fieldwork organization.

Whereas the estimated bias of proportion and means are not noteworthy, we find effects on correlations and regressions in sample E where the share of fakes is higher than in the other samples. We could show that some cheating interviewers are swamped with multivariate statistics and failed to reproduce the covariance between schooling and gross income as well as the linear regression on the log income. The selected suspicious records have a smaller impact on results than clearly faked interviews. Our empirical results show that the consequent parameters can be seriously biased. But the selected suspicious records have a smaller impact on results than clearly faked interviews.

Therefore we find empirical evidence for the finding by Schnell (1991), based on his simulation results, that even small proportion of fake interviews are an important problem in multivariate survey statistics.

References

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

Biemer, P. and Stokes, S. 1989. The Optimal Design Quality Control Samples to Detect Interviewer Cheating. *Journal of Official Statistics*, 5(1):23–39.

Boyle, J. 1994. An Application of Fourier Series to the Most Significant Digit Problem. *American Mathematical Monthly*, 101:879–976.

Cantwell, P. J., Bushery, J. M., and Biemer, P. 1992. Toward a Quality Improvement System for Field Interviewing: Putting Contant Reinterview Into Perspective. *Proceedings of the American Statistical Association (Survey Research Methods Section)*, pages 74–83.

Crespi, L. 1945. The Cheater Problem in Polling. *Public Opinion Quarterly*, Winter:431–445.

Diekmann, A. 2002. Diagnose von Fehlerquellen und methodische Qualität in der sozialwissenschaftlichen Forschung. Manuskript 06/2002, Institut für Technikfolgenabschätzung (ITA). Wien.

Epanechnikov, V. 1969. Nonparametric estimation of a multidimensional probability density. *Teoriya Veroyatnostej i Ee Primeneniya*, 14:156–162.

Evans, F. B. 1961. On Interviewer Cheating. *Public Opinion Quarterly*, 25:126–127.

Hamming, R. 1970. On the distribution of numbers. *Bell System Technical Journal*, 49:1609–1625.

Hill, T. P. 1995. A Statistical Derivation of the Significant-Digit Law. *Statistical Science*, 10:354–362.

Hood, C. C. and Bushery, J. M. 1997. Getting more Bang from the Reinterview Buck: Identifying "At Risk"" Interviewers. *Proceedings of the American Statistical Association (Survey Research Methods Section)*, pages 820–824.

Knuth, D. 1981. *The Art of Computer Programming 2: Seminumerical Programming.* Addison-Wesly, Reading, MA.

Koch, A. 1995. Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994. *ZUMA Nachrichten*, 36:89–105.

Moore, J. C. and Marquis, K. 1996. The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation. Working-Paper No. 212, U.S. Department of Commerce, Bureau of the Census.

Nigrini, M. 1999. I've got your number. *Journal of Accountancy*, 187:79–83.

Pinkham, R. 1961. On the distribution of the first significant digits. *The Annals of Mathematical Statistics*, 32:1223–1230.

Posch, P. N. 2003. Ziffernanalyse in der Fälschungsaufspürung. Das Benford-Phänomen und Steuererklärungen in Theorie und Praxis. Arbeitspapier vom Okt. 2003, Abteilung Finanzwirtschaft, University Ulm.

Reuband, K.-H. 1990. Interviews, die keine sind - "Erfolge" und "Mi"serfolge" beim Fälschen von Interviews. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 4:706–733.

Schnell, R. 1991. Der Einfluß gefälschter Interviews auf Survey-Ergebnisse. *Zeitschrift für Soziologie*, 20(1):25–35.

Schreiner, I., Pennie, K., and Newbrough, J. 491–496. Interviewer falsification in Census Bureau Surveys. *Proceedings of the American Statistical Association (Survey Research Methods Section).*

Scott, P. and Fasli, M. 2001. Benford's Law: An Empirical Investigation and a Novel Explanation. CSM Technical Report 349, Department of Computer Science, University Essex.

Stokes, L. S. and Jones, P. 696–198. Evaluation of the Interviewer Quality Control Procedure for the Post-Enumeration Survey. *Proceedings of the American Statistical Association (Survey Research Methods Section).*

Turner, C. F., Gribble, J. N., Al-Tayyib, A. A., and Chromy, J. R. 2002. Falsification in Epidemiological Surveys: Detection and Remediation (Prepublication Draft). Technical Papers on Health and Behavior Measurement, No. 53. Washington DC: Research Triangle Institute.