Diskussionspapierreihe
Working Paper Series

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

Applying "Benford's law" to the Crosswise Model:

Findings from an online survey on tax evasion

THORBEN KUNDT

Nr./ No. 148
JULY 2014

Department of Economics
Fächergruppe Volkswirtschaftslehre

Autoren / Authors

Thorben Kundt
Helmut Schmidt Universität Hamburg
Department of Economics
Holstenhofweg 85, 22043 Hamburg
Germany
kundt@hsu-hh.de

Redaktion / Editors

Helmut Schmidt Universität Hamburg / Helmut Schmidt University Hamburg
Fächergruppe Volkswirtschaftslehre / Department of Economics

Eine elektronische Version des Diskussionspapiers ist auf folgender Internetseite zu finden/
An electronic version of the paper may be downloaded from the homepage:
http://fgvwl.hsu-hh.de/wp-vwl

Koordinator / Coordinator

Klaus B. Beckmann
wp-vwl@hsu-hh.de

Helmut Schmidt Universität Hamburg / Helmut Schmidt University Hamburg
Fächergruppe Volkswirtschaftslehre / Department of Economics

Diskussionspapier Nr. 148
Working Paper No. 148

Applying "Benford's law" to the Crosswise Model: Findings from an online survey on tax evasion

THORBEN KUNDT

Zusammenfassung/ Abstract

Many surveys on sensitive topics such as tax evasion suffer from the reluctance of respondents to provide truthful answers which can cause downward-biased estimates. This paper addresses this problem by making use of a recent survey method (Crosswise Model) designed to provide positive incentives for respondents to answer sensitive questions more truthful. We extend the Crosswise Model by applying the so-called "Benford Illusion" which allows us to increase the precision of the Crosswise Model that is less statistically efficient than other methods. To test the effectiveness of the model in providing privacy protection, we carried out an online survey in which the respondents were randomly allocated into two splits differing only by the questioning technique applied. Our results suggest that the Crosswise Model can help to increase privacy protection compared to a simple direct questioning approach. As a consequence, survey estimates of tax evasion using the Crosswise Model are likely to become more valid. At the same time, we show that were able to obtain an efficient estimator without substantially decreasing privacy protection, even for a relatively small sample size.

JEL-Klassifikation / JEL-Classification: C83; H26

Schlagworte /Keywords: tax evasion; survey methodology; Crosswise Model; Benford's law

1. Introduction

The intense use of large-scale surveys such as the *World Values Survey* in the empirical tax compliance literature (see e.g. Torgler, 2007) emphasizes the role that surveys play as a relatively easy accessible means for researchers and policy makers for studying the anatomy of tax compliance. A major challenge for the design and implementation of surveys dealing with sensitive topics such as tax evasion, however, is to create positive incentives for respondents to answer these kind of questions truthfully.

There are various reasons why survey respondents might understate the true extent of tax evasion or even completely refrain from answering questions: some might have the feeling that their answers could be disclosed to third parties such as public authorities, others might overstate their tax honesty for reasons of social desirability, and some others might simply feel that questions on tax evasion are too intrusive (or, of course, a combination of these reasons) (Tourangeau and Yan, 2007). Consequently, survey estimates of tax evasion may be downward-biased which has serious implications for the assessment of costs and benefits of different policy measures taken to counter tax evasion (Kundt et al., 2013).

In order to cope with respondent dishonesty, a number of questioning strategies have been developed over the last decades. These techniques usually aim at increasing the (perceived) anonymity of questions on sensitive topics, thereby creating positive incentives for respondents to give true answers. Most prominently, various studies have applied (different variants of) the *Randomized Response Technique* (RRT)[1] (Warner, 1965). To increase privacy protection, RRT combines sensitive items with non-sensitive, usually unrelated ones for which the researcher knows the distribution of answers in advance. A randomization device determines which question is to be answered. Of course, the interviewer is unaware of the result of the randomization process and cannot directly observe which question the respondents answered. Yet, it is still possible to estimate the share of respondents for whom the sensitive question applied. However, a central drawback of the RRT is that it still leaves the respondents the opportunity to choose a self-protective strategy by ignoring the RRT mechanism and simply answering “no” in any case (Jann et al., 2012).

To overcome this problem, this study employs a recent survey method, the *Crosswise Model* (CM) (Yu et al., 2008). Comparable to RRT, the interviewer asks two questions under CM, a sensitive one with an unknown distribution, and a non-sensitive one with a known distribution. CM provides privacy protection because the respondents are offered two options to *jointly* answer both questions: (A) “yes” or “no” on both questions, or (B) a different answer on both questions. Thus, the interviewer can make no inferences about the respondent’s answer on the individual level. Yet it is possible to estimate the prevalence of the sensitive item for the sample as a whole.

CM seems better suited to study sensitive topics compared to RRT because it is designed such that a “no”-bias is ruled out (Jann et al., 2012). The only option not to reveal sensitive information is to refuse answering, but this option also applies for other questioning techniques (Kundt et al., 2013). Furthermore, CM does not rely on any kind of physical or digital randomization device. Jann et al. (2012) and Höglinger et al. (2014) argue that it is not easy to implement a suitable randomization procedure in online surveys and make respondents trust in it. A drawback of CM, however, is that the unrelated question adds noise to the data and the estimator becomes less efficient (Jann et al., 2012). For this reason, we apply a low-variance version of CM in our study.

[1] For a meta-study, see Lensvelt-Mulders et al. (2005); for an application to tax evasion, see Himmelfarb and Lickteig (1982), Houston and Tran (2001), and Musch et al. (2001).

Despite its potential to elicit more truthful answers, there are only few studies that have empirically applied CM so far. The thematic foci of these studies varied, covering *tax evasion* at the firm level (Kundt et al., 2013), *plagiarism and cheating* among students (Jann et al., 2012, Höglinger et al. 2014), and *illicit drug use* of students (Shamsipour et al., 2014), all of which representing presumably sensitive topics for the respondents. To assess the effectiveness of CM in providing privacy protection, the studies compared it to more conventional questioning methods such as asking respondents directly (*direct questioning, DQ*) which deliver no additional individual privacy protection. Throughout the surveys, CM systematically yielded higher prevalence estimates for the sensitive topics than the benchmark approaches; given that survey respondents are likely underreport social undesirable behavior, it follows that CM elicited more truthful answers ("more-is-better-assumption", Tourangeau and Yan, 2007).

This paper presents the results of a recent online survey on tax evasion that was designed to assess the benefits of studying tax evasion by means of CM. The study contributes to the literature as follows. First of all, there is only one study so far that implemented CM in an online survey (Höglinger et al., 2014) but which did not focus on tax evasion. Secondly, we are (to our knowledge) the first to focus on individual-level tax evasion using CM. Thirdly, we apply a low-variance version of CM by using a non-sensitive item that follows a logarithmic *Newcomb-Benford* distribution. By doing so, we are able to make use of the so-called "Benford Illusion" (Diekmann, 2012) which was successfully introduced for RRT, but not yet used for CM.[2]

We show that the prevalence of tax evasion for CM is significantly higher compared to a direct questioning approach. Multivariate analysis reveals that this result is robust. Utilizing the "Benford Illusion", we find that the estimator variance for CM remains fairly low. Moreover, our results suggest that CM significantly reduces item nonresponse, and that the "Benford Illusion" applies. The remainder of the paper is organized as follows. Section 2 introduces the Crosswise Model and discusses the application of the "Benford Illusion". Section 3 presents the research design and the hypothesis. Section 4 presents the results that are discussed in the concluding Section 5.

2. The Crosswise Model

2.1. Properties

The Crosswise Model is a recent approach suggested by Yu et al. (2008) aiming to reduce biased answers in surveys on sensitive topics[3]. Under the CM-design, the respondent is asked a non-sensitive question in addition to the sensitive one. The prevalence of the non-sensitive characteristic Y is known in advance and given by $p = \Pr(Y = 1)$. The probability that a respondent shares the sensitive characteristic X, on the other hand, cannot be observed and is noted with $\pi = \Pr(X = 1)$.

Unlike the RRT, CM "bundles" both the sensitive and the non-sensitive question by offering the respondent the following two options for a *joint* answer (Jann et al., 2012):

(A) Yes to both questions, or no to both questions
(B) Yes to one of the questions, and no to the other one

Option (A) (*same answer*) can be described with $\{X = 0 \cap Y = 0\} \cup \{X = 1 \cap Y = 1\}$ and option (B) (*different answer*) with $\{X = 1 \cap Y = 0\} \cup \{X = 1 \cap Y = 0\}$, respectively. The probabil-

[2] Combining Benford's Law and the Crosswise Model was suggested by Mark Trappmann at the 2012 FAU Workshop on Tax Compliance.

[3] The derivation of CM is largely based on Yu et al. (2008).

ity that a respondent opts for option (A) is given by $\lambda = (1-p)(1-\pi) + p\pi$. For option (B), the probability is $1-\lambda = p(1-\pi) + (1-p)\pi$, respectively.

Both options do not reveal anything definite about the respondent's true answers; with a larger *p*, the chances that the respondent answered "yes" to the non-sensitive question increases, or in other words, privacy protection rises with *p*. Because both questions are jointly answered, the respondent is unable to choose a self-protective strategy by simply saying "no" regardless of his true answer (Jann et al., 2012).

The unbiased maximum-likelihood estimate for the prevalence of the sensitive characteristic is given by (see Appendix):

$$\hat{\pi} = \frac{(\hat{\lambda} + p - 1)}{(2p-1)} \tag{1}$$

with $p \neq 0.5$, and $\hat{\lambda}$ being the observed proportion of respondents picking option (A).

The estimator variance is formally identical to the original RRT model as proposed by Warner (1965) (see Appendix):

$$Var(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)(2p-1)^2} = \underbrace{\frac{\hat{\pi}(1-\hat{\pi})}{(n-1)}}_{Sampling} + \underbrace{\frac{p(1-p)}{(n-1)(2p-1)^2}}_{Non-sensitive\ question} \tag{2}$$

From the right hand side of equation (2) can be seen that the estimator variance can be decomposed into a sampling part, and an additional source of error resulting form the introduction of the non-sensitive question. Hence, CM is statistically less efficient than other questioning methods. One straightforward way to deal with this loss of efficiency is to increase the sample size *n*. Yet, depending on the mode of data-collection, large samples are not always feasible. For face-to-face interviews, for example, recruiting and interviewing a large number of respondents can become very expensive.

A second way to improve the precision of CM estimator is to choose a non-sensitive item with a low prevalence (Jann et al., 2012) because the CM-variance is positively related with *p*, as demonstrated on the left hand side of equation (2). Lowering the level of *p*, on the other hand, comes at the expense of a decrease in privacy protection. This trade-off between efficiency and privacy protection is a drawback of CM (and some variants of RRT). However, in the next subsection we will argue that it is possible to obtain an efficient estimator while at the same time keeping privacy protection on a reasonable level. Analogous to the suggestion made by Moriarty and Wieseman (1976) the basic idea is to make respondents believe that it is more likely to ob-serve the non-sensitive item than actual data suggests. As Diekmann (2012) proposes, this discrepancy between perceived probability (p^*) and objective probability (*p*) could be ob-served for non-sensitive items that obey *Benford's law of the leading digits.*

2.2. Applying Benford's law

Benford's law of the leading digits goes back Benford (1938) who observed that the leading digits *d* = 1,2,3,…,9 in many, seemingly unrelated datasets follow a common logarithmic distribution which was first described by Newcomb (1988):

$$\Pr(d) = \log_{10}\left(1 + \frac{1}{d}\right) \tag{3}$$

Equation (3) shows that the probability to observe, for instance, the first digit 1 in a dataset that obeys Benford's law is 30.1% and thus much higher than for a uniform distribution. By contrast, the probability decreases to 4.6% at the rightmost tail of the distribution, that is, for the leading digit 9 (see also Section 3).

Although it has been claimed to be “a mathematic curiosity with no apparent useful application“ (Bolton and Hand, 2002, p.237) in the past, Benford’s law has received growing attention among social scientists over the last years. Table 1 displays some examples for recent applications. A particular area of interest was the detection of fraudulent or corrupted data. Based on the observation that human choices do not tend to be random (Nigrini, 1996), datasets created “by hand” should systematically deviate from Benford’s law (Hill, 1999).

Table 1: Recent applications of Benford’s law of the leading digits

Study	Type of data	Summary
Nigrini (1996)	Individual taxpayer data	The violations of Benford’s Law found in official taxpayer records are assumed to be a result of unplanned tax evasion (item overdeduction and income underdeclaration).
Diekmann (2007), Günnel and Tödter (2009), Tödter (2009)	Regression coefficients and standard in original research articles	The leading digits of regression coefficients and standard errors roughly approximate Benford’s law. Deviations could be a result of manipulated data.
Giles (2007)	Prices of eBay auctions	The first digits of pro-football ticket prices on eBay follow Benford’s Law. Violations of this pattern can be interpreted as a hint for collusion among bidders.
Mir (2012)	Country-level adherents of different world religions	Observed total count of countries for which the number of a religions’ adherents starts with 1,2,3,…,9. Except from Christianity, all major world religions follow Benford’s Law.
Diekmann (2012)	Swiss house-numbers	The leading digits of Swiss house-numbers obey Benford’s law. The perceived distribution among survey respondents, however, is more close to uniform (“Benford Illusion”).

More important for our study, Benford’s law has recently been used in survey research to improve the efficiency of the Randomized Response Technique. Analogous to CM, some common variants suffers of RRT suffer from a trade-off between statistical precision on the one hand, and anonymity on the other hand (see Section 2.1). The RRT-model proposed by Diekmann (2012) was modified such that the distribution of the non-sensitive item followed Benford’s law. In particular, respondents in a survey on plagiarism were asked about the leading digit of a friend’s house-number. Benford (1938) observed that leading digits of house-numbers were a typical example for a logarithmic distribution similar to equation (3); the data presented by Diekmann (2012) (for Switzerland) and in Section 3 of this paper (for Germany) confirm this observation.

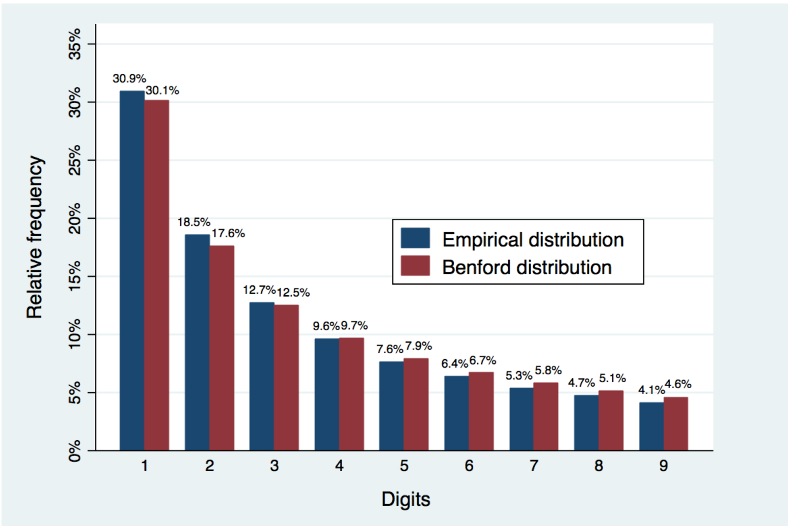
Diekmann (2012) assumed that respondents were likely to systematically make false assumptions about the prevalence of the leading digits of house-numbers (“*Benford Illusion*”). As outlined in Section 2.1, creating this kind of discrepancy between perceived and objective probabilities is a reasonable way to increase the efficiency of methods such as RRT and CM. The survey results presented by Diekmann (2012) suggested that the respondents’ subjective assessments of the probabilities significantly deviated from what could be observed from the data. Thus, the “Benford Illusion” successfully allowed for decreasing the RRT-variance (which depended on the objective probability p) without suffering from a loss in anonymity (which depended on the perceived p^*). By the same line of reasoning, we apply the “Benford Illusion” to CM in order to improve the statistical efficiency of the estimator, as we will demonstrate in the next section.

3. Research design and hypotheses

Our data comes from a short online survey (average time for responding in minutes: 5.49, *SD* = 6.32) on tax-related topics carried out in 2014. The survey was implemented using the commercial provider *Unipark*. Respondents were actively recruited online via university mailing lists and in business- as well as social networks. In an introductory statement, respondents were informed about the survey content. They were assured complete anonymity and that the data provided by them would be solely used for scientific purposes.

In the survey, interviewees were randomly allocated to one of two splits that solely differed by questioning method.[4] In the first split, respondents were asked directly about tax evasion (*direct questioning, DQ*); in the second split, we applied the Crosswise Model. Net of refusals, 137 respondents answered the question on tax evasion in the DQ condition, and 256 in CM condition. The CM-split was oversampled by approximately two (1.87) to further increase precision of the CM-estimator (see Jann et al., 2012). In the CM-split, the respondents were given additional information on how the method worked and then told to think of the first digit of a friend's house-number in Germany. They were encouraged to look for the address in their mobile phones or an address-book if they did not remember it by heart. We chose the leading digits of German house-numbers because they nearly perfectly obey Benfords's law (Figure 1). Another advantage of using house-numbers is that credible data (at least for Germany) is relatively easy to access and that respondents should be able to recall a friend's house-number.

Figure 1: Relative frequencies of the first digits of German house-numbers



<u>Notes:</u> *N* = 3,326,422; own calculations based on data collected via the open source project "OpenStreetMap"

After having finished the instructions, respondents had to answer the following non-sensitive question in conjunction with the sensitive one, with two response options (A) and (B) analogous to those described in Section 2.1:

"Is the first digit of your of your friend's house-number 7, 8, or 9?"

[4] At the very beginning of the survey, a uniformly distributed variable *c* with *c* = 1,2,3 was randomly generated for each respondent. If *c* = 1, respondents were allocated to the DQ-split, if *c* = 2 or *c* = 3, respondents were allocated to the CM-split, respectively.

We focused on the right tail of the distribution with the leading digits 7,8 and 9. From Figure 1 we can see that the chances to pick a German house-number starting with 7,8, or 9 are p = 14.1 percent. For the “Benford Illusion” to work, respondents have to believe that this figure is higher and closer to uniform. Based on this assumption, which is central for the “Benford Illusion”, we make to following hypothesis:

> *Hypothesis 1 (“Benford Illusion”)*: The mean perceived probability of observing a German house-number starting with 7,8, or 9 is larger than the objective probability (*Null-hypothesis:* $p^* \leq p$; *Alternative hypothesis:* $p^* > p$).

Question wording of the sensitive item was identical for both conditions and chosen such that it came close to the definition of tax evasion as provided by §370 of the German tax code (“Abgabenordnung”). Specifically, we asked the respondents the following question (translated from German):

> *“Have you ever intentionally underdeclared income and/or made false statements to the tax office in order to pay less or no income taxes?”*

As it is common practice in the literature, we compare CM to DQ serving as benchmark. DQ provides no additional privacy protection and allows the respondents to give a false negative answer. As a consequence, the share of dishonest answers is likely to be higher for the benchmark approach.

> *Hypothesis 2 („More is better assumption“)*: The prevalence estimate for tax evasion will be higher for CM (π_{CM}) than for DQ (π_{DQ}) (*Null-hypothesis* $\pi_{CM} \leq \pi_{DQ}$; *Alternative hypothesis:* $\pi_{CM} > \pi_{DQ}$).

4. Results

4.1. Sample description

In Table 2 we present summary statistics for the respondents’ socio-demographic background and the experience in paying income taxes (filed income tax return)[5] for the total sample and differentiated by questioning method (p-values for the differences between DQ and CM are reported in the rightmost column; *Null-hypothesis:* $Mean_{DQ} = Mean_{CM}$). None of the variables differ significantly between splits and we can infer that the prevalence estimates were driven by heterogeneous subgroups.

Table 2: Summary statistics

	Total		**DQ**		**CM**		
Variable	Mean	SD	Mean	SD	Mean	SD	*p-value*
Age	30.8	10.7	31.2	11.3	30.6	10.4	0.53
Male	0.66	0.47	0.69	0.46	0.62	0.48	0.43
German	0.98	0.15	0.99	0.12	0.97	0.17	0.40
Net-income (*EUR*)	2038.69	1666.86	2022.94	902.89	2117.66	1970.16	0.61
Self-employed	0.15	0.36	0.16	0.37	0.17	0.35	0.68
Filed income tax return *(one time or more)*	0.89	0.31	0.92	0.21	0.88	0.33	0.19

[5] Some of the respondents who did not file an income tax return might still have been subject to income taxation without explicitly mentioning it, for example if they had exclusively worked in the informal economy.

4.2. Perceived probabilities

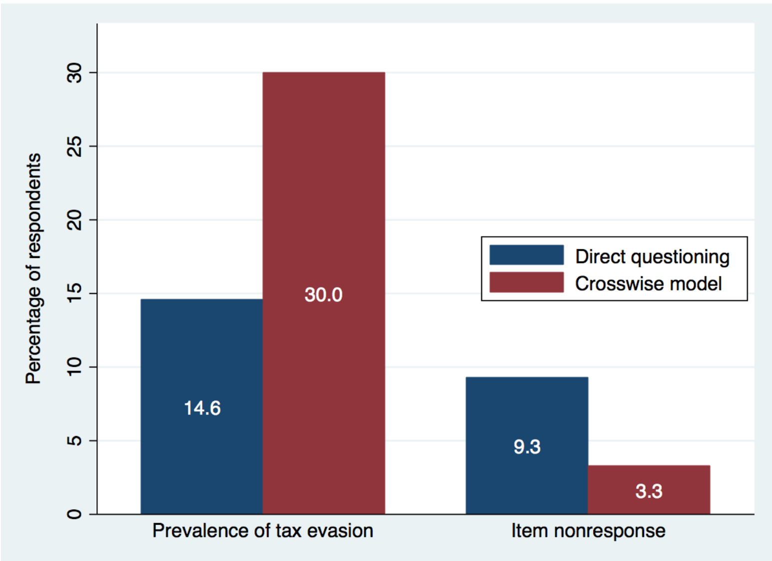
The central assumption of the "Benford Illusion" as applied in our design is that respondents systematically overestimate the prevalence of the non-sensitive item which on the one hand increases the perceived privacy protection of CM, and on the other side reduces the variance of the CM-estimator which depends on the objective probability to observe the non-sensitive item (*Hypothesis 1*). To evaluate the effectiveness of the "Benford Illusion", respondents were asked to estimate the probability for a German house-number to start with 7,8, or 9 after they had finished the crosswise questions. We explicitly informed the respondents *not* to recall of the house-number of their friends (as they did before), but to think of Germany in general in order not to raise suspicion about the anonymity provided by CM.

The mean perceived value chosen by the respondents was 0.27 (SD = 0.14) which turns out to be significantly larger than the empirically observed value of 0.14, $t(236) = 13.94$, $p < .01$ (one-tailed). This results supports our hypothesis of the "Benford Illusion" and consequently, we were able to increase precision of the CM-estimator while at the same time keeping anonymity up.

4.3. Prevalence estimates

Figure 2 displays the prevalence estimates for tax evasion (i.e., the percentage of positive answers) using direct questioning and the Crosswise Model.

Figure 2: Prevalence estimates of tax evasion and item nonresponse



In the DQ-sample, 14.6 (SE = 3.0) percent of the respondents admitted that they have cheated on their income taxes by making false statements to the tax authorities and/or underdeclaring income. The relatively higher level of anonymity provided by the online mode might explain this surprisingly large fraction (Musch et al., 2001). Additionally, we framed the question in a rather neutral way without explicitly mentioning the term tax *evasion*, which may have also increased the percentage of truthful answers (see Section 3).

Taking a look at the results of the CM-questions, we find the prevalence estimate of 30.0 (SE = 4.2) percent to exceed the result for direct questioning. The difference of 15.4 percentage points between the two questioning methods is highly statistically significant, $z = 3.00$, $p < .01$ (one-tailed). Following Cohen's (1988) criterion, the effect size of $d = 4.14$ further emphasizes the relevance of this result. Taken together, we can confirm *Hypothesis 2* which stated that CM would increase anonymity and deliver higher point estimates for the sensitive question on tax evasion ("more-is-better assumption"). At the same time, the standard error of 4.2 percent remains comparably low.

Figure 2 also shows that item nonresponse was only 3% under CM. When being asked directly, 9% of the respondents refused to answer. A two-sample proportion test reveals that the share is significantly larger than for CM, z = -2.53, p < .01 (one-tailed), d = 3.27. Given that refusing to answer represents another respondent strategy not to reveal sensitive information, CM might be well suited to study sensitive behavior from this perspective.

4.4. Robustness of results

Complexity of CM

A critique of rather complex techniques such as RRT or CM is that they require a high level of cognitive effort and might lead to confusion among respondents, especially when they involve physical or digital randomization devices (Umesh and Peterson, 1991; Höglinger et al., 2014). To control for this, our questionnaire included a question that assessed whether respondents understood the CM-mechanism and that it protected their privacy. We did not explicitly ask respondents how much they *trusted* the technique because this might have raised suspicion, given that we assured privacy in the first place. Yet, 63.0 percent of the respondents fully understood CM-mechanism and that it protected their privacy; another 21.0 percent knew that CM delivered privacy protection without understanding the exact mechanism; the remaining 16.0 percent told us that they did not understand CM. However, restricting the sample to the first two categories of respondents only marginally changes the results (*prevalence estimate*: 31.7%, *SE* = 4.6%).

Multivariate analysis

The particular design of CM prevents us from using standard multivariate techniques to check the robustness of our results. Given that respondents jointly answered both sensitive and nosensitive items, we are unable to correctly assign the "yes" and "no"-answers to the respective questions (Jann et al., 2012). Yet, we can use a modified version of the standard logistic regression approach introduced by Jann (2011) (randomized response logit, *rrlogit*). Although originally designed to analyze data from RRT-models, rrlogit is also applicable to CM. The rrlogit-model differs from the regular logistic regression approach because it allows for including the probability for a positive answer on the non-sensitive item that is known in advance. The dependent variable is defined as a response variable that takes the value one if the respondent provided the same answer on both CM-questions, and zero otherwise.

A particular advantage of the model is that we can jointly analyze data from both splits by using observation-specific probabilities because asking respondents directly represents a special case of CM with no additional privacy protection (Jann et al., 2012): if we assume that $p = 0$, then the only possible outcome for option (A) (same answer) is "no to both questions" and (A) reduces to $\lambda = (1 - \pi)$; likewise, we can directly infer from option (B) (different answer) that the respondent's answer on the sensitive question was "yes" $(\lambda = \pi)$.

The results of the multivariate analysis are presented in Table 3. In addition to sociodemographics, the models include a dummy variable that indicates whether the respondent has been self-employed, and the self-reported knowledge of the German tax system (*Tax knowledge*, 1 = Very poor; 6 = Very good). The models (2) and (3) also feature questions particularly related to CM and are restricted to a sub-sample of respondents (CM-split only). As noted above, participants were asked whether they understood CM which is covered by the variable *Understood CM*. The variable *Perceived probability* indicates how participants estimated the chances to pick a German house-number starting with 7,8, or 9. Furthermore, we included a dummy (*Benford Illusion*) that takes on the value zero if the estimated probability was below or equal to the observed probability of 14.1%, and one otherwise.

Table 3 : Results from the randomized response logit

	Model (1)		Model (2)		Model (3)	
Variable	*Coef.*	*SE*	*Coef.*	*SE*	*Coef.*	*SE*
Split (*ref.: DQ*)	0.96^{***}	0.33				
Socio-demographics						
Age	-0.00	0.02	-0.00	0.02	-0.00	0.02
Gender (*ref.: Male*)	0.13	0.36	-0.02	0.49	0.04	0.48
Net-income	0.00	0.00	-0.00	0.00	-0.00	0.00
Self-employed (*ref.: No*)	-0.16	0.47	-0.48	0.59	-0.48	0.59
Tax knowledge	-0.02	0.01	-0.02	0.02	-0.02	0.02
Understood CM (*ref.: No*)						
Partially			0.28	0.76	0.28	0.76
Fully			0.10	0.65	0.11	0.65
Perceived probability			0.01	0.02		
Benford Illusion (*ref.: p* < 0.141*)					0.11	0.60
Constant	-1.93^{**}	1.02	0.19	1.32	0.28	1.34
No. observations	365		225		225	
Pseudo R^2	0.03		0.01		0.01	

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Most coefficients turned out to be insignificant because of the low statistical power of the randomized response logit (Jann et al., 2012). However, the split-dummy is highly significant and positive which confirms our hypothesis that CM is better suited to study tax evasion than asking directly, given that the "more-is-better" assumption holds. Furthermore, the variables related to CM in models (2) and (3) are in the direction that we expected. There is a positive effect on tax evasion when the respondents (either partially or fully) understood the method and that it protected their privacy compared to those who did not. We also find a positive effect of the perceived probability and the dummy capturing the "Benford Illusion". Yet, these effects remain insignificant due to the lack of statistical power of the rrlogit.

5. Discussion and conclusions

This paper implemented a recent questioning technique (Crosswise Model, CM) to study self-reported tax evasion in an online survey. CM seems well-suited to study sensitive topics such as tax evasion because it increases the (perceived) level of question anonymity and may thus lead to more truthful answers among respondents. One particular advantage of CM over other methods such as RRT is that it prevents the respondents from answering "no" on sensitive questions even if the true answer is positive by combining sensitive and non-sensitive items which have to be answered jointly. Our research design differed from previous surveys because we applied a non-sensitive item for which the known distribution of answers obeyed "Benford's Law" of the leading digits. We chose this particular design to make use of the "Benford Illusion" in order to let participants believe that the prevalence of the non-sensitive item (in our case: the first digits 7,8, and 9 of German house-numbers) was higher than actual data tells us. By doing so, we could obtain an efficient estimator for CM without decreasing the level of anonymity for the sensitive item.

We tested the effectiveness of CM by randomly allocating respondents into two splits, one asking survey participants directly about tax evasion, and the second one applying CM. Our focus was on tax evasion because questions on this topic are likely to provoke biased answers and item non-response. As a central result of this study, we showed that the prevalence estimate for tax evasion using CM was significantly higher than for the benchmark approach. If the "more is better"-assumption holds, than CM should have provided a higher level of anonymity, leading to more valid results. In addition, CM also helped to reduce item nonresponse. This finding complements the results of Shamsipour et al. (2014) who showed that applying CM in a survey on drug use decreased item nonresponse compared to asking directly. Furthermore, although being relatively complex, most of the respondents understood the mechanics of CM and, even more importantly, that it protected their privacy.

On average, respondents in our survey thought that the probability to observe a German house-number that starts with a 7, 8, or 9 was 27%. However, using a large-scale dataset with more than 3 million observations, we showed that the probability is only 14% and, what is also important, that the first digits nearly perfectly fit a Newcomb-Benford distribution. We could thus confirm the observation of the "Benford Illusion" which was successfully implemented for RRT by Diekmann (2012) and Höglinger et al. (2014). The "Benford Illusion" helped us to reduce the standard error of the Crosswise-estimator that remained low even for our relatively small sample.

Of course, there are some limitations for our study. With respect to external validity, our results should be treated with caution because our sample is not necessarily representative for German taxpayers. However, the majority of the respondents could at least be characterized as experienced income-taxpayers. Furthermore, because this study was primarily intended to assess the effectiveness of (a low-variance version) of CM, our main task was to ensure that the splits did not vary with respect to certain important respondent characteristics. We showed that the splits did not significantly differ in this respect, and we can rule out the possibility that our results were driven by heterogeneous treatments. Upcoming surveys might apply CM to a representative and maybe larger sample of respondents, if the purpose is to generalize the results to the whole universe of (German) taxpayers.

From an economic perspective, the prevalence estimate of 30% for tax evasion seems to be substantial. However, our sample primarily consisted of wage earners who usually have fewer possibilities to evade income taxes than self-employed. Moreover, although CM offered a high level of privacy protection, some of the respondents might still have not been convinced by the technique, causing them to provide wrong answers. Taken together, the estimated proportion of 30% might be interpreted as a lower bound.

Finally, as outlined by Kundt et al. (2013), CM only allows for studying the *prevalence* of tax evasion or other types of sensitive behavior by using dichotomous questions (i.e., "yes" or "no"). Policy makers might be also interested in the overall scope of revenues foregone. To gather quantitative data it would be necessary to ask a whole set of CM questions (Kundt et al., 2013) or to apply a modified version of CM.

Acknowledgements

The author would like to thank Christoph Bühren, Andreas Diekmann and Ben Jann for useful comments.

References

Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, *78*, 551-72.

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, 235-249.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale: Lawrence Erlbaum Associates.

Diekmann, A. (2007). Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *Journal of Applied Statistics*, *34*(3), 321-329.

Diekmann, A. (2012). Making Use of "Benford's Law" for the Randomized Response Technique. *Sociological Methods & Research*, *41*(2), 325–334.

Giles, D. E. (2007). Benford's law and naturally occurring prices in certain ebaY auctions. *Applied Economics Letters*, *14*(3), 157-161.

Günnel, S., & Tödter, K. H. (2009). Does Benford's Law hold in economic research and forecasting? *Empirica*, *36*(3), 273-292.

Hill, T. P. (1999). The difficulty of faking data. *Chance*, *12*(3), 27-31.

Himmelfarb, S., & Lickteig, C. (1982). Social desirability and the randomized response technique. *Journal of Personality and Social Psychology*, *43*(4), 710–717.

Höglinger, M., Jann, B., & Diekmann, A. (2014). Sensitive Questions in Online Surveys: An Experimental Evaluation of the Randomized Response Technique and the Crosswise Model. *University of Bern Social Sciences Working Paper No. 9*.

Houston, J., & Tran, A. (2001). A survey of tax evasion using the randomized response technique. *Advances in Taxation*, *13*, 69–94.

Jann, B. (2011). RRLOGIT: Stata module to estimate logistic regression for randomized response data. *Statistical Software Components*.

Jann, B., Jerke, J., & Krumpal, I. (2012). Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism. *Public Opinion Quarterly*, *76*(1), 32–49.

Kundt, T. C., Misch, F., & Nerré, B. (2013). Re-assessing the merits of measuring tax evasions through surveys: Evidence from Serbian firms. *ZEW Discussion Papers No. 13-047*.

Lensvelt-Mulders, G. J., Hox, J. J., Van der Heijden, P. G., & Maas, C. J. (2005). Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation. *Sociological Methods & Research*, *33*(3), 319–348.

Mir, T. A. (2012). The law of the leading digits and the world religions. *Physica A: Statistical Mechanics and its Applications*, *391*(3), 792-798.

Moriarty, M., & Wiseman, F. (1976). On the choice of a randomization technique with the randomized response model. In *Proceedings of the Social Statistics Section, American Statistical Association*, 624-626.

Musch, J., Bröder, A., & Klauer, K. C. (2001). Improving Survey Research on the World-Wide Web Using the Randomized Response Technique. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science* (pp. 179-192). Lengerich, Germany: Pabst Science Publishers.

Newcomb, S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics*, *4*, 39-40.

Nigrini, M. (1996). A taxpayer compliance application of Benford's law. *Journal of the American Taxation Association*, *18*(1), 72-92.

Shamsipour, M., Yunesian, M., Fotouhi, A., Jann, B., Rahimi-Movaghar, A., Asghari, F., & Akhlaghi, A. A. (2014). Estimating the Prevalence of Illicit Drug Use Among Students Using the Crosswise Model. *Substance Use & Misuse*, *Early Online,* 1-8.

Torgler, B. (2007). Tax compliance and tax morale: a theoretical and empirical analysis. Cheltenham: Edward Elgar Publishing.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859-883.

Tödter, K. H. (2009). Benford's Law as an Indicator of Fraud in Economics. *German Economic Review*, *10*(3), 339-351.

Umesh, U. N., & Peterson, R. A. (1991). A Critical Evaluation of the Randomized Response Method Applications, Validation, and Research Agenda. *Sociological Methods & Research*, *20*(1), 104-138.

Warner, S. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*(309), 63–69.

Yu, J.-W., Tian, G.-L., & Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika*, *67*(3), 251–263.

Appendix: Derivation of the Crosswise Model

The appendix is largely based on Warner (1965) because both the CM-estimator and –variance are formally identical to the original RRT-model.

In the first step, we recall the unobserved probabilities for options (A) and (B) which are given by $\lambda = (1-p)(1-\pi) + p\pi$ and $1-\lambda = p(1-\pi) + \pi(1-p)$, respectively. If n_1 respondents chose option (A), and the remaining respondents (n - n_1) opted for option (B), we can formulate the following log-likelihood function:

$$L = [(1-\pi)(1-p) + \pi p]^{n_1} [\pi(1-p) + p(1-\pi)]^{n-n_1} \quad \text{(A1)}$$

$$\log L = n_1 \log[(1-\pi)(1-p) + \pi p] + (n-n_1)\log[\pi(1-p) + p(1-\pi)] \quad \text{(A2)}$$

The first order condition is given by:

$$\frac{n_1(2p-1)}{\pi p(1-\pi)(1-p)} = \frac{(n-n_1)(2p-1)}{\pi(1-p)p(1-\pi)} \Leftrightarrow (1-\pi)(1-p) + \pi p = \frac{n_1}{n} \quad \text{(A3)}$$

With $\hat{\lambda} = n_1/n$ representing the maximum likelihood estimate for λ, solving for π gives the unbiased estimator $\hat{\pi}$:

$$\hat{\pi} = \frac{(\hat{\lambda} + p - 1)}{(2p-1)},\ p \neq 0.5 \quad \text{(A4)}$$

Because $n_1 \sim binomial(n, \hat{\lambda})$, with $E(n_1) = n\hat{\lambda}$ and $Var(n_1) = n\hat{\lambda}(1-\hat{\lambda})$, the variance is derived as follows (Yu et al., 2008; Kundt et al., 2013):

$$Var(\hat{\pi}) = Var\left(\frac{(\hat{\lambda} + p - 1)}{(2p-1)}\right) = \frac{n\hat{\lambda}(1-\hat{\lambda})}{n^2(2p-1)} = \frac{\hat{\lambda}(1-\hat{\lambda})}{n(2p-1)} \quad \text{(A5)}$$

Finally, correcting (A5) by $n/(n-1)$ (Bessel's correction) gives:

$$Var(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)(2p-1)^2} = \underbrace{\frac{\hat{\pi}(1-\hat{\pi})}{(n-1)}}_{Sampling} + \underbrace{\frac{p(1-p)}{(n-1)(2p-1)^2}}_{Non-sensitive\ question} \quad \text{(A6)}$$

As (A6) demonstrates, we can disaggregate the variance into a sampling part, and an additional part which results from the introduction of the non-sensitive question.

Die komplette Liste der Diskussionspapiere ist auf der Internetseite veröffentlicht / for full list of papers see:
http://fgvwl.hsu-hh.de/wp-vwl

2014

147 Beckmann, Klaus; Reimer, Lennart: Dynamiken in asymmetrischen Konflikten: eine Simulationsstudie

146 Herzer, Dierk: Unions and income inequality: a heterogeneous panel cointegration and causality analysis, July 2014

145 Beckmann, Klaus; Franz, Nele; Schneider, Andrea: Intensive Labour Supply: a Menu Choice Revealed Preference Approach for German Females and Males, June 2014

144 Beckmann, Klaus; Franz, Nele; Schneider, Andrea: On optimal tax differences between heterogenous groups, May 2014

143 Berlemann, Michael; Enkelmann, Sören: Institutions, experiences and inflation aversion, May 2014

142 Beckmann, Klaus; Gattke, Susan: Tax evasion and cognitive dissonance, April 2014

141 Herzer, Dierk; Nunnenkamp, Peter: Income inequality and health – evidence from developed and developing countries, April 2014

140 Dewenter, Ralf; Heimeshoff, Ulrich: Do Expert Reviews Really Drive Demand? Evidence from a German Car Magazine, March 2014

139 Dewenter, Ralf; Heimeshoff, Ulrich: Media Bias and Advertising: Evidence from a German Car Magazine, March 2014.

138 Beckmann, Klaus; Reimer, Lennart: Dynamics of military conflict from an economics perspective, February 2014.

2013

137 Christmann, Robin: Tipping the Scales - Conciliation, Appeal and the Relevance of Judicial Ambition.

136 Hessler, Markus; Loebert, Ina: Zu Risiken und Nebenwirkungen des Erneuerbare-Energien-Gesetzes, June 2013.

135 Wesselhöft, Jan-Erik: The Effect of Public Capital on Aggregate Output- Empirical Evidence for 22 OECD Countries -, June 2013.

134 Emrich, Eike; Pierdzioch, Christian; Rullang, Christian: Zwischen Ermessensfreiheit und diskretionären Spielräumen: Die Finanzierung des bundesdeutschen Spitzensports – eine Wiederholungsstudie, April 2013.

133 Christmann, Robin: Vertragliche Anreize und die Fehlbarkeit des Richters – Der ungewisse Gang vor Gericht und sein Einfluss auf eine Verhaltenssteuerung im BGB-Vertragsrecht, March 2013.

132 Gerrits, Carsten: Internetnutzer und Korruptionswahrnehmung - Eine ökonometrische Untersuchung, February 2013.

131 Freese, Julia: The regional pattern of the U.S. house price bubble - An application of SPC to city level data, January 2013.

2012

130 Kruse, Jörn: Unabhängige staatliche Institutionen: Funktionalität und demokratische Legitimation, November 2012.

129 Andrae, Jannis: Unabhängigkeit von Institutionen - Gründe bzw. Ursachen und Kriterien zur Beurteilung der Unabhängigkeit von öffentlichen Institutionen im demokratischen Rechtsstaat, November 2012.

128 Andrae, Jannis: Ideengeschichtliche Aspekte unabhängiger Institutionen, November 2012.

127 Pfeiffer, Christoph P.: Causalities and casualties: Media attention and terrorism, 1970–2010, November 2012.

126 Pierdzioch, Christian; Emrich, Eike: A Note on the International Coordination of Anti-Doping Policies, November 2012.

125 Berlemann, Michael; Wesselhöft, Jan-Erik: Estimating Aggregate Capital Stocks Using the Perpetual Inventory Method – New Empirical Evidence for 103 Countries –, October 2012.

124 Berlemann, Michael; Freese, Julia; Knoth, Sven: Eyes Wide Shut? The U.S. House Market Bubble through the Lense of Statistical Process Control, October 2012.

123 Pierdzioch, Christian; Emrich, Eike; Klein, Markus: Die optimierende Diktatur – Politische Stabilisierung durch staatlich verordnetes Doping am Beispiel der DDR, August 2012.

122 Flatau, Jens; Emrich, Eike; Pierdzioch, Christian: Zum zeitlichen Umfang ehrenamtlichen Engagements in Sportvereinen – sozioökonomische Modellbildung und empirische Prüfung, August 2012.

121 Pfeiffer, Christoph P.: The curse of anxiety-pleasure: Terrorism, the media, and advertising in a two-sided market framework, August 2012.

120 Pitsch, Werner; Emrich, Eike; Pierdzioch, Christian: Match Fixing im deutschen Fußball: Eine empirische Analyse mittels der Randomized-Response-Technik, August 2012.