Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663

www.czasopisma.uni.lodz.pl/foe/
1(358) 2022
DOI: https://doi.org/10.18778/0208-6018.358.01

Barbara Kowalczyk https://orcid.org/0000-0002-5407-3438

SGH Warsaw School of Economics, Collegium of Economic Analysis
Institute of Econometrics, Mathematical Statistics Unit, Warsaw, Poland, bkowal@sgh.waw.pl

An Analysis of the Properties of a Newly Proposed Non-Randomised Response Technique

Abstract: Non-randomised response (NRR) techniques are modern and constantly evolving methods intended for dealing with sensitive topics in surveys, such as tax evasion, black market, corruption etc. The paper introduces a new NRR technique that can be seen as a generalisation of the well-known crosswise model (CM). In the paper, methodology of the new generalised crosswise model (GCM) is presented and the maximum likelihood estimator of the unknown population sensitive proportion is obtained. Also, the problem of privacy protection is discussed. The properties of the newly proposed GCM are examined. Then the GCM is compared with the traditional CM. The paper shows that mathematically the CM is a special case of the newly proposed generalised CM and that this generalisation has high practical relevance.

Keywords: indirect questioning, sensitive questions, non-randomised response techniques, crosswise model, ML estimation, degree of privacy protection

JEL: C13, C18, C83

1. Introduction

In many economic and social studies, researchers are interested in features, attributes and characteristics that are inherently sensitive. Questions about sensitive topics include tax frauds, corruption, illegal work, black market, abortion, politically incorrect views, drug use, and many more.

The first statistical method of indirect questioning intended for dealing with sensitive features was proposed by Warner (1965). The method is called mirrored question design. Nowadays it is one of many techniques belonging to a larger group of methods named randomised response techniques (RRTs). The common feature of all RRTs is the use of a randomising device by respondents. A randomising device can be a coin, dice, spinner, deck of cards or a random number generator. Other elemental randomised response techniques are: unrelated question design (Greenberg, Abul-Ela, Horvitz, 1969) and forced response design (Boruch, 1971; Fox, Tracy, 1986). A very good summary of basic RRTs can be found in Blair, Imai and Zhou (2015).

Randomised response techniques developed very strongly in the 21st century. Theoretical developments include much more complicated models (Abdelfatah, Mazloum, 2015; Batool, Shabbir, Hussain, 2017), optional randomised response techniques (Khalil, Zhang, Gupta, 2021) and scrambled randomised responses (Vishwakarma, Singh, 2021). Applications of RRTs include research on inappropriate sexual behaviour (Rueda, Cobo, López-Torrecillas, 2020), illegal waste disposal (Chong et al., 2019), ethics in business (Chu, So, Chung, 2018), corruption (Gingerich, 2010), and many more.

Although RRTs are well established in statistical theory and practice, there is an ongoing debate as to whether these methods really help to elicit truthful answers to questions that are inherently sensitive (Coutts, Jann, 2011; Wolter, Preisendörfer, 2013; Leslie et al., 2018). The biggest disadvantage of randomised response techniques is the necessity to use a randomising device by respondents. Randomised response techniques usually cannot be used in telephone or internet surveys. To avoid this drawback, Yu, Tian, and Tang (2008) introduced crosswise and triangular models that resemble RRTs but do not need any randomising device. The new methods received the name of non-randomised response (NRR) techniques and gained a great deal of attention in the 21st century. Groenitz (2014) proposed a new NRR technique for a categorical sensitive variable called the diagonal model, Tian (2014) proposed the parallel model, Arnab, Shangodoyin, and Arcos (2019) extended the parallel model to complex survey designs. Applications of non-randomised response techniques include research on xenophobia (Hoffman, Meisters, Musch, 2020), mental stress among students (Erdmann, 2019), attitudes towards female leaders (Hoffmann, Musch, 2019), attitudes towards Muslims (Johann, Thomas, 2017), premarital sex among university students (Wu, Tang, 2016), and tax evasion (Korndörfer, Krumpal, Schmukle, 2014).

Section 2 of the paper presents the statistical background of the crosswise model proposed by Yu, Tian, and Tang (2008). In Section 3, the methodology of the newly proposed generalised crosswise model (GCM) is presented, the maximum likelihood estimator of the population sensitive proportion is obtained, and the degree of privacy protection is given. In Section 4, the properties of the new GCM are discussed and comparisons with other NRR models are made. The practical relevance of the newly proposed generalised crosswise model is shown. The article ends with Conclusions set out in Section 5.

2. Non-randomised response techniques

To allow for further examinations and comparisons, we briefly present key results for the crosswise model pioneered by Yu, Tian, and Tang (2008). Although the paper introduces a generalisation of the crosswise model, to present a broader point of view, we will also refer to the triangular model given again in Yu, Tian, and Tang (2008). In the mentioned models, the researcher is interested in a sensitive variable Z with binary outcomes, i.e. $Z \sim Bernoulli(\pi)$. The parameter under study is an unknown population sensitive proportion π.

Crosswise model (Yu, Tian, Tang, 2008)

In the crosswise model, a neutral question with binary outcomes and a known distribution is used, i.e. we have $Q \sim Bernoulli(q)$, where q is a known population proportion such that $q \neq 0.5$. It is assumed that neutral and sensitive questions are unrelated, i.e. that Q and Z are independent. Respondents are presented with two questions simultaneously, one neutral and one sensitive. They are instructed to report only if answers to those two questions are the same.

The observable variable in this model is Y, where:

$$Y = \begin{cases} 1 \text{ if two answers are YES or two answers are NO} \\ 0 \quad \text{otherwise} \end{cases}. \tag{1}$$

We have:

$$P_{CM}(Y=1) = q\pi + (1-q)(1-\pi) \tag{2}$$

and

$$\hat{\pi}_{CM} = \frac{\overline{y} - 1 + q}{2q - 1} \tag{3}$$

$$Var\left(\hat{\pi}_{CM}\right)=\frac{\pi\left(1-\pi\right)}{n}+\frac{q\left(1-q\right)}{n\left(2q-1\right)^{2}}=\frac{\left[q\pi+\left(1-q\right)\left(1-\pi\right)\right]\left[1-q\pi-\left(1-q\right)\left(1-\pi\right)\right]}{n\left(2q-1\right)^{2}} \quad (4)$$

A popular indicator of respondents' privacy in indirect methods of questioning is the degree of privacy protection (DPP) defined as $DPP = P(Z = 1|Y)$. For the CM, we have:

$$P_{CM}\left(Z=1\,|\,Y=0\right)=\frac{\left(1-q\right)\pi}{\left(1-q\right)\pi+q\left(1-\pi\right)}. \quad (5)$$

$$P_{CM}\left(Z=1\,|\,Y=1\right)=\frac{q\pi}{q\pi+\left(1-q\right)\left(1-\pi\right)}. \quad (6)$$

It has to be emphasised that the choice of model parameters should be based on some sort of compromise between the efficiency of estimation and the degree of privacy protection. For example, in the crosswise model, $Var\left(\hat{\pi}_{TM}\right)$ for $q\in\left(0;0.5\right)$ is an increasing function of q, therefore, small values of q are preferred. At the same time, $P_{CM}(Z = 1|Y = 0)$ is a decreasing function of q, therefore, large values of q are needed. Consequently, in research on sensitive topics both the efficiency of estimation and respondents' privacy need to be taken into account simultaneously. This discussion will be continued in the next sections of the paper.

3. New proposed generalised crosswise model

In this section, we introduce a generalised crosswise model (GCM). We present the methodology of this new technique, obtain the maximum likelihood estimator of the unknown population sensitive proportion and derive the degree of privacy protection. Then we show that the crosswise model proposed by Yu, Tian, and Tang (2008) is statistically a special case of the GCM proposed here.

For better understanding and also for further comparisons, we first present a version of this new technique for two neutral questions only. Then the model will be further generalised.

We are interested in a sensitive variable Z with binary outcomes, i.e. $Z \sim Bernoulli(\pi)$. The parameter under study is an unknown population sensitive proportion π. We use two neutral questions with binary outcomes and a known distribution, i.e. we have $Q \sim Bernoulli(q)$, $R \sim Bernoulli(r)$, where q and r are known population proportions and $gr \neq (1 - q)(1 - r)$. We assume that three questions are unrelated, i.e. mathematically, we assume that Q, R, Z are mutually independent.

Respondents are presented with all three questions simultaneously, i.e. they are presented with one sensitive and two neutral questions at the same time. Then they are asked only about the three joined questions. In particular, they are instructed to report if answers to all three questions are the same, i.e. if all answers are YES or all answers are NO. They are not asked about individual questions. Therefore, none of the variables Q, R, Z is directly observable (they are latent variables). The only variable observable in this model is Y, where:

$$Y = \begin{cases} 1 & \text{if three answers are YES or three answers are NO} \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

We have:

$$P_{GCM}(Y=1) = qr\pi + (1-q)(1-r)(1-\pi). \tag{8}$$

Now we construct a log-likelihood function:

$$\begin{aligned} \ln L(\pi, q, r; y_1, \ldots, y_n) = n\bar{y}\ln\left[qr\pi + (1-q)(1-r)(1-\pi)\right] + \\ + n(1-\bar{y})\ln\left[1 - qr\pi - (1-q)(1-r)(1-\pi)\right]. \end{aligned} \tag{9}$$

By using standard derivation, we obtain the maximum likelihood estimator of the unknown sensitive proportion:

$$\hat{\pi}_{GCM} = \frac{\bar{y} - (1-q)(1-r)}{qr - (1-q)(1-r)} \tag{10}$$

The variance of this estimator is given by:

$$Var(\hat{\pi}_{GCM}) = \frac{\left[qr\pi + (1-q)(1-r)(1-\pi)\right]\left[1 - qr\pi - (1-q)(1-r)(1-\pi)\right]}{n(r+q-1)^2}. \tag{11}$$

To analyse the properties of the new technique, let us derive the degree of privacy protection (DPP), which is a popular indicator of privacy in indirect methods of questioning. The DPP is defined as $DPP = P(Z = 1|Y)$, and it describes the probability of possessing a sensitive attribute given the answer Y. For the newly proposed model, we obtain:

$$P_{GCM}(Z=1 \mid Y=0) = \frac{(1-qr)\pi}{(1-qr)\pi + (1-(1-q))(1-r)(1-\pi)}. \tag{12}$$

$$P_{GCM}(Z=1 \mid Y=1) = \frac{qr\pi}{\text{qr}\pi + (1-q)(1-r)(1-\pi)}. \tag{13}$$

Now let us generalise this model to a larger number of neutral variables. We use k, $k \geq 1$ neutral questions with binary outcomes and a known distribution, i.e. we have $Q_i \sim Bernoulli(q_i)$, where q_i are known population proportions, $i = 1, \ldots, k$ and $\prod_{i=1}^{k} q_i \neq \prod_{i=1}^{k}(1-q_i)$. We assume that all questions are unrelated, i.e. that $Q_1, \ldots, Q_k, Z$ are mutually independent.

Respondents are presented with all $(k + 1)$ questions simultaneously, i.e. they are presented with one sensitive and k neutral questions at the same time. Then they are instructed to report if answers to all $(k + 1)$ questions are the same, i.e. if all answers are YES or all answers are NO. The only variable observable in this model is Y, where:

$$Y = \begin{cases} 1 & \text{if answers to all } (k+1) \text{ questions are the same} \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

We have:

$$P_{GCM}(Y=1) = \pi \prod_{i=1}^{k} q_i + (1-\pi) \prod_{i=1}^{k}(1-q_i). \tag{15}$$

The log-likelihood function is:

$$\begin{aligned} \ln L(\pi, q_1, \ldots, q_n; y_1, \ldots, y_n) = n\bar{y} \ln\left[\pi \prod_{i=1}^{k} q_i + (1-\pi)\prod_{i=1}^{k}(1-q_i)\right] + \\ +(n - n\bar{y}) \ln\left[1 - \pi \prod_{i=1}^{k} q_i - (1-\pi)\prod_{i=1}^{k}(1-q_i)\right]. \end{aligned} \tag{16}$$

The maximum likelihood estimator of the unknown sensitive proportion is:

$$\hat{\pi}_{GCM} = \frac{\bar{y} - \prod_{i=1}^{k}(1-q_i)}{\prod_{i=1}^{k} q_i - \prod_{i=1}^{k}(1-q_i)} \tag{17}$$

The variance of this estimator is given by:

$$Var(\hat{\pi}_{GCM}) = \frac{\left[\pi \prod_{i=1}^{k} q_i + (1-\pi)\prod_{i=1}^{k}(1-q_i)\right]\left[1 - \pi \prod_{i=1}^{k} q_i - (1-\pi)\prod_{i=1}^{k}(1-q_i)\right]}{n\left(\prod_{i=1}^{k} q_i - \prod_{i=1}^{k}(1-q_i)\right)^2}. \tag{18}$$

For the proposed model, we obtain:

$$P_{GCM}(Z=1 \mid Y=0) = \frac{\left(1 - \prod_{i=1}^{k} q_i\right)\pi}{\left(1 - \prod_{i=1}^{k} q_i\right)\pi + \left(1 - \prod_{i=1}^{k}(1-q_i)\right)(1-\pi)}. \tag{19}$$

$$P_{GCM}\left(Z=1|Y=1\right)=\frac{\pi\prod_{i=1}^{k}q_i}{\pi\prod_{i=1}^{k}q_i+\left(1-\pi\right)\prod_{i=1}^{k}\left(1-q_i\right)}. \quad (20)$$

It can be easily seen that the GCM simplifies to the crosswise model proposed by Yu, Tian, and Tang (2008) for $k = 1$. And formulas (14–20) simplify to formulas (1–6). Therefore, the crosswise model is a special case of the new GCM. From a mathematical point of view, it is obvious that we have obtained a more general model. Now an important question arises as to whether such a theoretical generalisation has any practical relevance. The answer is ‘yes’ and it will be justified it in the next section.

4. Properties of the generalised crosswise model

In real surveys, we usually cannot choose values of q_i freely, as we are limited by the kind of questions we can ask in a given population. Consequently, having some questions at our disposal, we have given the values of q_i, $i = 1, \ldots, k$ in advance. Therefore, it will be convenient to compare the traditional CM and the newly introduced GCM for selected model parameters, i.e. for selected values of q_i, $i = 1, \ldots, k$ and π.

First, let us study the efficiency of the GCM in comparison with the CM introduced by Yu, Tian, and Tang (2008). For this purpose, let us define:

$$Eff\left(\hat{\pi}_{GCM}\right)=\frac{Var\left(\hat{\pi}_{GCM}\right)}{Var\left(\hat{\pi}_{CM}\right)}, \quad (21)$$

where $Var\left(\hat{\pi}_{GCM}\right)$ is given in formula (18) and $Var\left(\hat{\pi}_{CM}\right)$ is given in formula (4). In both models, we assume the same values of q_1 and π. This means that in both models we consider the same sensitive question and one neutral question. The only difference is that in the generalised crosswise model succeeding neutral questions are added. Table 1 presents the efficiency of the GCM for selected model parameters and different values of neutral questions (different values of k).

Table 1. Efficiency $Eff\left(\hat{\pi}_{GCM}\right)$ of the proposed GCM for selected model parameters

	Sensitive population proportion π						
Parameters q_i	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
			***k* = 2**				
$q_1 = q_2 = 0.2$	1.266	1.147	1.054	0.974	0.898	0.818	0.728
$q_1 = q_2 = 0.3$	1.103	1.027	0.957	0.890	0.824	0.755	0.681

	Sensitive population proportion π						
Parameters q_i	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
q_1 =0.2; q_2 = 0.1	0.869	0.822	0.779	0.735	0.685	0.626	0.552
q_1 = 0.2; q_2 = 0.3	1.870	1.644	1.479	1.346	1.232	1.124	1.014
***k* = 3**							
$q_1 = q_2 = q_3 = 0.3$	1.531	1.370	1.227	1.095	0.966	0.838	0.704
$q_1 = q_2 = q_3 = 0.4$	1.141	1.062	0.985	0.909	0.834	0.758	0.680
q_1 = 0.3; q_2 = 0.3; q_3 = 0.2	1.167	1.052	0.946	0.846	0.746	0.642	0.532
q_1 = 0.3; q_2 =0.3; q_3 = 0.4	2.102	1.874	1.674	1.494	1.325	1.159	0.992
***k* = 4**							
$q_1 = q_2 = q_3 = q_4 = 0.3$	2.250	1.971	1.726	1.503	1.292	1.086	0.875
$q_1 = q_2 = q_3 = q_4 = 0.45$	1.247	1.177	1.108	1.039	0.970	0.902	0.833
$q_1 = q_2 = q_3$ =0.4; q_4 = 0.3	1.121	1.018	0.916	0.816	0.716	0.615	0.511
$q_1 = q_2 = q_3$ = 0.4; q_4 = 0.45	1.981	1.817	1.657	1.503	1.350	1.199	1.047

Source: own calculations

It can be seen in Table 1 that using more than one neutral question can be beneficial as far as the efficiency of the estimation is concerned. In particular, in the cases of model parameters corresponding to the shaded cells in Table 1, the newly introduced GCM is more efficient than the traditional CM.

Now that we know that adding neutral questions can be beneficial, an important question arises. Namely, is there an optimal number of neutral questions? Or equivalently, is there an optimal number of k? The answer is no. It depends on the unknown sensitive proportion π. To show that, let us examine the following example. Let us say that we have at our disposal four different neutral questions, answers to which are Bernoulli distributed with the same parameter $q = 0.3$. Shall we use all four questions, three questions, two questions, or only one question? There is no single answer to this question. If the real sensitive proportion is $\pi = 0.1$, then

$$Var_{k=1}\left(\hat{\pi}_{GCM}\right) \le Var_{k=2}\left(\hat{\pi}_{GCM}\right) \le Var_{k=3}\left(\hat{\pi}_{GCM}\right) \le Var_{k=4}\left(\hat{\pi}_{GCM}\right).$$

Therefore, in this particular case, it is best to use one neutral question only. But if the real sensitive proportion is $\pi = 0.3$, then we have

$$Var_{k=2}\left(\hat{\pi}_{GCM}\right) \le Var_{k=1}\left(\hat{\pi}_{GCM}\right) \le Var_{k=3}\left(\hat{\pi}_{GCM}\right) \le Var_{k=4}\left(\hat{\pi}_{GCM}\right).$$

Therefore, in this particular case, it is best to use two neutral questions. Last but not least, if $\pi = 0.8$, then

$$Var_{k=3}\left(\hat{\pi}_{GCM}\right) \le Var_{k=2}\left(\hat{\pi}_{GCM}\right) \le Var_{k=4}\left(\hat{\pi}_{GCM}\right) \le Var_{k=1}\left(\hat{\pi}_{GCM}\right).$$

Therefore, in this particular case, it is best to use three neutral questions.

Let us now consider the problem of privacy protection. For respondents to feel safe, the DPP should be small. On the other hand, the extreme version of *DPPP* = 0 frees respondents from any suspicion of possessing sensitive attributes. Let us note that for the triangular model (Yu, Tian, Tang, 2008), it holds:

$$P(Z=1\,|\,Y=0)=0.$$

This means that by answering‛ No,' i.e. by answering that $Y = 0$, in this model, respondents can protect themselves from being accused of possessing a sensitive attribute. This is not a desirable feature of the model due to the fact that the model possesses an option for full protection. Respondents may be tempted by this option and may select it to show that they do not have the undesired characteristic. For the crosswise model (Yu, Tian, Tang, 2008) and for the new proposed GCM, it holds:

$$DPP=P(Z=1\,|\,Y=i)\neq 0 \text{ for } i=0,1.$$

Thus, the CM and GCM do not have any option for full protection, which is considered an advantage. Therefore, in this regard, both the CM and GCM models are superior to the triangular model.

Therefore, let us now examine in detail the degree of privacy protection for both the GCM and CM for the same model parameters as in Table 1. Tables 2 and 3 present ratios of degrees of privacy protection defined as:

$$r_{GCM}(Z=1\,|\,Y=i)=\frac{P_{GCM}(Z=1\,|\,Y=i)}{P_{CM}(Z=1\,|\,Y=i)} \text{ for } i=0,1, \tag{22}$$

where $P_{GCM}(Z = 1|Y = i)$ is given in formulas (19) and (20) and $P_{CM}(Z = 1|Y = i)$ is given in formulas (5) and (6).

Table 2. Ratios of degree of privacy protection $r_{GCM}(Z = 1|Y = 0)$

	Sensitive population proportion π						
Parameters q_i	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
***k* = 2**							
$q_1 = q_2 = 0.2$	0.743	0.800	0.844	0.880	0.909	0.933	0.954
$q_1 = q_2 = 0.3$	0.804	0.837	0.867	0.893	0.915	0.936	0.954
$q_1 =0.2$; $q_2 = 0.1$	0.910	0.933	0.950	0.963	0.972	0.980	0.986
$q_1 = 0.2$; $q_2 = 0.3$	0.623	0.696	0.757	0.808	0.851	0.889	0.922
***k* = 3**							
$q_1 = q_2 = q_3 = 0.3$	0.686	0.733	0.777	0.816	0.853	0.887	0.918

	Sensitive population proportion π						
Parameters q_i	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
$q_1 = q_2 = q_3 = 0.4$	0.820	0.843	0.865	0.886	0.907	0.927	0.946
$q_1 = 0.3$; $q_2 = 0.3$; $q_3 = 0.2$	0.739	0.781	0.818	0.852	0.882	0.910	0.935
$q_1 = 0.3$; $q_2 =0.3$; $q_3 = 0.4$	0.640	0.691	0.738	0.783	0.825	0.864	0.901
***k* = 4**							
$q_1 = q_2 = q_3 = q_4 = 0.3$	0.615	0.668	0.717	0.764	0.809	0.851	0.891
$q_1 = q_2 = q_3 = q_4 = 0.45$	0.878	0.892	0.906	0.920	0.934	0.947	0.961
$q_1 = q_2 = q_3 =0.4$; $q_4 = 0.3$	0.796	0.822	0.846	0.870	0.893	0.916	0.938
$q_1 = q_2 = q_3 = 0.4$; $q_4 = 0.45$	0.764	0.792	0.820	0.847	0.874	0.900	0.926

Source: own calculations

Table 3. Ratios of degree of privacy protection $r_{GCM}(Z = 1|Y = 0)$

	Sensitive population proportion π						
Parameters q_i	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
***k* = 2**							
$q_1 = q_2 = 0.2$	0.255	0.262	0.270	0.280	0.294	0.314	0.345
$q_1 = q_2 = 0.3$	0.440	0.454	0.470	0.491	0.517	0.552	0.600
$q_1 =0.2$; $q_2 = 0.1$	0.114	0.117	0.122	0.127	0.135	0.147	0.165
$q_1 = 0.2$; $q_2 = 0.3$	0.435	0.443	0.454	0.467	0.484	0.508	0.543
***k* = 3**							
$q_1 = q_2 = q_3 = 0.3$	0.191	0.199	0.210	0.224	0.243	0.270	0.310
$q_1 = q_2 = q_3 = 0.4$	0.462	0.483	0.507	0.536	0.571	0.615	0.672
$q_1 = 0.3$; $q_2 = 0.3$; $q_3 = 0.2$	0.112	0.117	0.124	0.134	0.146	0.165	0.194
$q_1 = 0.3$; $q_2 =0.3$; $q_3 = 0.4$	0.295	0.307	0.321	0.340	0.364	0.397	0.444
***k* = 4**							
$q_1 = q_2 = q_3 = q_4 = 0.3$	0.082	0.086	0.092	0.099	0.109	0.123	0.146
$q_1 = q_2 = q_3 = q_4 = 0.45$	0.569	0.593	0.621	0.652	0.688	0.730	0.779
$q_1 = q_2 = q_3 =0.4$; $q_4 = 0.3$	0.202	0.215	0.232	0.254	0.282	0.320	0.376
$q_1 = q_2 = q_3 = 0.4$; $q_4 = 0.45$	0.380	0.400	0.424	0.452	0.488	0.533	0.594

Source: own calculations

As we can see in Tables 2 and 3, for all model parameters considered in these tables (shaded cells), the generalised crosswise model increased respondents' privacy, i.e. decreased conditional probabilities $P(Z = 1|Y = 0)$ and $P(Z = 1|Y = 1)$ that respondents possess a sensitive attribute. Moreover, in many cases, this improvement is extremely significant. Therefore, the new GCM can find applications especially in surveys where sensitivity of the studied variable is exceptionally high. It should be noted, however, that this is not always the case. There are exceptions to this rule. For example, for $k = 2$; $q_1 = 0.4$; $q_1 = 0.2$, we have $r_{GCM}(Z = 1|Y = 0) > 1$ for all values of the sensitive proportion π.

In Tables 2–3, we have indicated situations in which it is superior to the CM in terms of privacy protection. We have also emphasised that both the CM and GCM do not have any option for full protection, which is considered an advantage over the non-randomised triangular model.

Last but not least, let us note that the new GCM increases simultaneously both the efficiency of estimation and respondents' privacy for all model parameters corresponding to the shaded cells in Table 1 compared to the traditional non-randomised CM. This proves that the presented generalisation is very important from a practical point of view.

5. Conclusions

Non-randomised response techniques are innovative and valuable tools for dealing with sensitive questions in surveys. Existing NRR models provide many possibilities for applications. Nevertheless, it is still valuable to develop existing theory. In the paper, a generalisation of the existing crosswise model has been introduced. It has been shown that the classical crosswise model pioneered by Yu, Tian, and Tang (2008) is a special case of the obtained generalised crosswise model. Additionally, in the paper, it has been indicated that this generalisation is also relevant from a practical point of view. For selected model parameters, the introduced GCM increases not only the efficiency of estimation but also respondents' privacy as compared to the traditional CM. Such an improvement is very important as far as sensitive topics are concerned.

References

Abdelfatah S., Mazloum R. (2015), *Efficient estimation in a two-stage randomized response model*, "Mathematical Population Studies", vol. 22, pp. 234–251.

Arnab R., Shangodoyin D.K., Arcos A. (2019), *Nonrandomized Response model for Complex Survey Designs*, "Statistics in Transition New Series", vol. 20, pp. 67–86.

Batool F., Shabbir J., Hussain Z. (2017), *An improved binary randomized response model using six decks of cards*, "Communications in Statistics – Simulation and Computation", vol. 46, pp. 2548–2562.

Blair G., Imai K., Zhou Y.-Y. (2015), *Design and Analysis of the Randomized Response Technique*, "Journal of the American Statistical Association", vol. 110, pp. 1304–1319.

Boruch R.F. (1971), *Assuring Confidentiality of Responses in Social Research: A Note on Strategies*, "American Sociologist", vol. 6, pp. 308–311.

Chong A.C.Y., Chu A.M.Y, So M.K.P, Chung R.S.W. (2019), *Asking Sensitive Questions Using the Randomized Response Approach in Public Health Research: An Empirical Study on the Factors of Illegal Waste Disposal*, "International Journal of Environmental Research and Public Health", vol. 16(6), https://www.researchgate.net/publication/331848508_Asking_Sensitive_Questions_Using_the_Randomized_Response_Approach_in_Public_Health_Research_An_Empirical_Study_on_the_Factors_of_Illegal_Waste_Disposal (accessed: 4.01.2022).

Chu A.M.Y., So M.K.P., Chung R.S.W. (2018), *Applying the Randomized Response Technique in Business Ethics Research: The Misuse of Information Systems Resources in the Workplace*, "Journal of Business Ethics", vol. 151, pp. 195–212.

Coutts E., Jann B. (2011), *Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT)*, "Sociological Methods & Research", vol. 40, pp. 169–193.

Erdmann A. (2019), *Non-Randomized Response Models: An Experimental Application of the Triangular Model as an Indirect Questioning Method for Sensitive Topics*, "Methods, Data, Analyses", vol. 13, pp. 139–167.

Fox J.A., Tracy P.E. (1986), *Randomised Response: A Method for Sensitive Surveys*, Sage Publications, Beverly Hills.

Gingerich D.W. (2010), *Understanding Off-The-Books Politics: Conducting Inference on the Determinants of Sensitive Behavior with Randomized Response Surveys*, "Political Analysis", vol. 18, pp. 349–380.

Greenberg B.G., Abul-Ela A.-L.A., Horvitz D.G. (1969), *The Unrelated Question Randomized Response Model: Theoretical Framework*, "Journal of the American Statistical Association", vol. 64, pp. 520–539.

Groenitz H. (2014), *A new privacy-protecting survey design for multichotomous sensitive variables*, "Metrika", vol. 77, pp. 211–224.

Hoffmann A., Musch J. (2019), *Prejudice against women leaders: Insights from an indirect questioning approach*, "Sex Roles", vol. 80, pp. 681–692.

Hoffmann A., Meisters J., Musch J. (2020), *On the validity of non-randomized response techniques: an experimental comparison of the crosswise model and the triangular model*, "Behavior Research", vol. 52, pp. 1768–1782.

Johann D., Thomas K. (2017), *Testing the Validity of the Crosswise Model: A Study on Attitudes Towards Muslims Survey Methods: Insights from the Field*, https://surveyinsights.org/?p=8887 (accessed: 4.01.2022).

Khalil S., Zhang Q., Gupta S. (2021), *Mean estimation of sensitive variables under measurement errors using optional RRT models*, "Communications in Statistics – Simulation and Computation", vol. 50, pp. 1417–1426.

Korndörfer M., Krumpal I., Schmukle S. C. (2014), *Measuring and explaining tax evasion: Improving self-reports using the crosswise model*, "Journal of Economic Psychology", vol. 45, pp. 18–32.

Leslie K.J., Loewenstein G.F., Acquisti A., Vosgerau J. (2018), *When and Why Randomized Response Techniques (Fail To) Elicit the Truth*, "Organizational Behavior and Human Decision Processes", vol. 148(C), pp. 101–123.

Rueda M.M., Cobo B., López-Torrecillas F. (2020), *Measuring Inappropriate Sexual Behavior Among University Students: Using the Randomized Response Technique to Enhance Self-Reporting*, "Sex Abuse", vol. 32, pp. 320–334.

Tian G.L. (2014), *A new non-randomized response model: The parallel model*, "Statistica Neerlandica", vol. 68, pp. 293–323.

Vishwakarma G.K., Singh N. (2021), *Computing the effect of measurement errors under additive scramble response of the sensitive variable*, "Journal of Computational and Applied Mathematics", vol. 395, 113593.

Warner S.L. (1965), *Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias*, "Journal of the American Statistical Association", vol. 60, pp. 63–69.

Wolter F., Preisendörfer P. (2013), *Asking sensitive questions: An evaluation of the randomized response technique versus direct questioning using individual validation data*, "Sociological Methods & Research", vol. 42, pp. 321–353.

Wu Q., Tang M.L. (2016), *Non-randomized response model for sensitive survey with noncompliance*, "Statistical Methods in Medical Research", vol. 25, pp. 2827–2839.

Yu J.W., Tian G.L., Tang M.L. (2008), *Two new models for survey sampling with sensitive characteristic: Design and analysis*, "Metrika", vol. 67, pp. 251–263.

Analiza własności nowo zaproponowanej techniki nierandomizowanych odpowiedzi

Streszczenie: Techniki nierandomizowanych odpowiedzi to nowoczesne i stale rozwijające się metody przeznaczone do radzenia sobie z tematami drażliwymi, takimi jak oszustwa podatkowe, czarny rynek, korupcja itp. W artykule zaproponowano nową technikę nierandomizowanych odpowiedzi, którą można traktować jako uogólnienie znanego modelu krzyżowego. Przedstawiono metodykę nowego uogólnionego modelu krzyżowego oraz podano estymator największej wiarygodności dla nieznanej populacyjnej frakcji cechy drażliwej. Omówiono również problem ochrony prywatności. Przeanalizowano własności nowo zaproponowanego modelu, a następnie porównano go z tradycyjnym modelem krzyżowym. Pokazano, że klasyczny model krzyżowy jest specjalnym przypadkiem zaproponowanego modelu uogólnionego. Wykazano również, że to uogólnienie ma duże znaczenie dla praktyki.

Słowa kluczowe: ankietowanie pośrednie, pytania drażliwe, techniki nierandomizowanych odpowiedzi, model krzyżowy, estymacja NW, stopień ochrony prywatności

JEL: C13, C18, C83

Received: 2021-08-08; verified: 2021-11-04. Accepted: 2022-02-08