STATISTICS IN TRANSITION new series, June 2021
Vol. 22, No. 2, pp. 15–39, DOI 10.21307/stattrans-2021-015
Received – 16.06.2020; accepted – 23.07.2020

Estimating completion and breakoff functions in a business web survey

Berislav Žmuk[1]

ABSTRACT

A business web survey should be of an appropriate length. On the one hand it should include all the questions which are important to the researcher, but on the other hand, it should not be too long as the breakoff rate in this case tends to be high, resulting in a low response rate. In consequence, the researcher is forced to invest more time and money in order to reach a sample size which would enable an appropriately performed statistical analysis. In this paper, completion and breakoff times are observed and compared across different questionnaire and respondent characteristics. A regression modelling approach has been adopted to estimate the completion and breakoff functions to help a researcher determine which respondents completed a questionnaire and which broke it off too quickly or too slowly. By omitting such respondents, a researcher is able to obtain the relevant estimates more efficiently. In addition, the completion and breakoff functions offer a better insight into the completion and breakoff development rates, allowing the researcher to make a better-informed decision as to whether the survey requires any modifications or not.

Key words: breakoff function, business web survey, completion function, questionnaire.
JEL: C12, C20, C83.

1. Introduction

The most common way of conducting a research in the modern business world is through a survey. However, conducting a survey is not so each task as it may look at the first sight. Namely, there are always more questions which researchers want to include into the questionnaire. On the other hand, business or enterprises do not have unlimited time and resources to be spent on taking part in a survey. In addition, if the survey questionnaire is too long, respondents in enterprises could simply break off at a certain point. In this way, the research would not get all answers from such respondents. It has to be emphasized that the survey questionnaire could be considered

[1] University of Zagreb, Faculty of Economics and Business, Department of Statistics, Trg J. F. Kennedyja 6, HR-10000 Zagreb, Croatia. E-mail: bzmuk@efzg.hr. ORCID: https://orcid.org/0000-0003-3487-1376.

too long not only because it has too many questions, but because it has too difficult questions for the respondents as well.

Consequently, the question is how to make a survey of an optimal length. As indicators of an optimal survey length, a researcher could observe completion and breakoff times. If the completion time is too long, a research should reduce the number of questions and lower the level of their complexity. If the breakoff time is too short, the same should be done. However, when the completion time is observed it is assumed that respondents answer all given questions, whereas when the breakoff time is observed the time when they stopped providing answers is observed. In this way, by observing breakoff times, some impact on overall response rates could be given as well.

There are three main outcomes related to the survey response of respondents. So, a respondent can make a decision not to participate in the survey. In that case, survey nonresponse appears (Kish, 1995, Groves, 2006). While survey nonresponse is something unwanted for a researcher, the most wanted situation for a researcher is when a respondent completes the survey by answering all survey questions. The third possible outcome is that a respondent starts with the survey but gives up at some point in the survey. In that case, one can speak about survey breakoff. There are some other types of breakoff like unit and item nonresponse (Peytchev, 2009) but due to limitation of the paper length it will not be discussed here. Also, because survey nonresponse is not the focus of this paper, it will not be observed in more details.

The goal of the paper is to introduce a way how completion and breakoff times could be used in the process of survey preparation. Consequently, in the paper completion and breakoff functions are going to be estimated. Some authors already tried to estimate impact of different items on completion and breakoffs in web surveys. However, they focused only on point estimation and comparison between different surveys. Crawford, Couper and Lamias (2001) estimated the likelihood of breakoff by sharing different information to respondents about the survey length. However, that information was not true for all respondents. Galesic (2006) used three surveys of different length in her analysis. It came out that the 30-minute survey had 40% higher breakoff risk than at a 10-minute survey. By reducing the length of the survey to 20 minutes, the breakoff risk was 20% higher than at a 10-minute survey. Göritz (2006) has shown that surveys with incentives have in average 27% higher probability of being complete. Peytchev (2011) compared breakoffs for respondents in consecutive surveys. He used multinomial regression approach to estimate odds ratios for different previous survey outcomes and for different characteristics of respondents (gender, race, years in school). Similarly, Blumenberg et al. (2018) also compared breakoff rates by comparing different characteristics of respondents by applying logistic regression. Vehovar and Cehovin (2014) conducted meta-analysis and found that 80% of all breakoffs occur on the first introductory pages. Mittereder (2019) used survival analysis for predicting

when respondents will breakoff from a web survey. The analysis is based on comparing survey results from two different years.

In order to be able to estimate the completion and breakoff functions, the paradata from a business web survey conducted on a sample of Croatian enterprises is used. Still, in addition to the research question whether the completion and breakoff functions potentially could be used for finding optimal survey time, another research question is whether they could be used in the process of detecting respondents who completed the questionnaire too fast (speeders) or too slowly. Accordingly, two research hypotheses have been defined. The first research hypothesis is that optimal survey lengths, estimated by the completion functions, are different for different questionnaire designs and for respondents of different characteristics. The second research hypothesis contains the assumption that the completion and breakoff functions can be used to detect too fast and too slow respondents.

The paper is organized as follows. After a brief introduction and an overview of previous research, data used in the analysis are presented. In addition, in the second chapter methods applied in the analysis are introduced as well. In the third chapter the completion times and the completion function are analysed, whereas in the fourth chapter focus is given to breakoff times and breakoff function estimation. Discussion is provided in the fifth chapter, whereas the sixth chapter concludes.

2. Data and methodology

Data for the analysis are taken from the business web survey which was conducted in 2016. The target population were active enterprises which were registered at the Commercial Court in Croatia until July 1, 2016 and which have a public available e-mail address. The responses were collected in the period from October 4 to December 31, December 2016. In that period, in addition to the initial invitation to participate in the survey research, a total of two reminders were sent.

The topic of the survey was statistics methods use in enterprises. The questionnaire mostly consisted of about 20 close-ended questions. However, some of the close-ended questions were binary questions, some questions were Likert scale based, and some were single select multiple choice questions. Also, a couple of open-ended questions were added mostly as a support to respondents if they wanted to emphasize something. According to that, this web survey could be observed as a medium long and complex questionnaire (1ka, 2020a,b, Žmuk, 2017).

For the purpose of the analysis in this paper, instead of responses, completion and breakoff times will be the focus. The completion time is defined here as the time which a respondent needed to fully complete the questionnaire. The time is started to be measured when a respondent opened the questionnaire and it is stopped when the

respondent selects to submit his answers. On the other hand, the breakoff time is measured from the point when a respondent opened the questionnaire to the point when the last action of the respondent in the questionnaire in the used survey software is registered. In order to collect data LimeSurvey software is used.

Initially different questionnaire versions were prepared. Different questionnaire versions were randomly associated to the population units. However, it has to be emphasized that different questionnaire versions were allocated to approximately the same number of the population units. Still, due to different response rates at different questionnaire versions, a different number of responses were collected in each questionnaire version.

The questionnaire versions were different according to the fact whether the pictures were included in the questionnaire or not and what kind of pictures were there. The first questionnaire version did not include any pictures, whereas other two questionnaire versions included pictures. In one, the so-called "positive", and in the second one "negative" pictures are implemented. Positive pictures included some positive information like picture with positive business trend or a table full of advanced statistical books. On the other hand, the questionnaire version with negative pictures included pictures where something "negative" is shown like a negative business trend or a table with small number of basic statistical books.

The questionnaire versions were different according to the number of questions presented to a respondent per questionnaire screen. In the first questionnaire version all questions were immediately shown to respondents. In the second questionnaire version questions were grouped into logical sections. The third questionnaire version presents only one question per questionnaire screen at a time.

The completion and breakoff times are observed according to the main characteristic of the participating enterprises as well. In this way, enterprises according to their legal form, size and main activity are inspected. According to the Enterprises Act, a distinction between joint stock enterprises, limited liability enterprises and simple limited liability enterprises is made (Narodne novine, 2011). Furthermore, according to the Accounting Act the enterprises are stratified by their size on small, medium and large enterprises (Narodne novine, 2015). Also, enterprises are observed according to their main activities. According to National Classification of Economic Activities (Narodne novine, 2007) overall 21 areas of enterprises activities are recognized. However, for the purpose of the analysis those areas are merged and, in this way, the number of main activities groups is reduced to four (industrial enterprises, trade enterprises, service enterprises, other enterprises).

The completion and breakoff times will be analysed separately but the same approaches and methods will be used in both cases. First of all, data cleaning will be performed. Namely, it is possible that some respondents needed too much time to

complete the questionnaire. For example, they started the survey but were distracted by a phone call, e-mails, other colleagues and similar. In this way, the completion time is longer that in reality was. Also, it is possible that the survey software wrongly measured times. For example, that the survey is completed in zero seconds. Such cases should also be omitted from the further analysis.

After the data is cleaned, descriptive statistics methods will be used to observe the completion and breakoff times. Because there are a lot of unique completion and breakoff times (measured in seconds), the descriptive statistics analysis will be applied on grouped data. In order to be able to compare results the same time groups will be used for completion and for breakoff times analysis.

The structure of respondents at different questionnaire versions will be compared by using the chi-square test for equality of three or more population proportion results. The null hypothesis of the chi-square test is that proportions are statistically equal across all observed populations. On the other hand, in the alternative hypothesis the assumption that not all proportions across the different populations are equal is incorporated. The empirical chi-square value is calculated as follows:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(m_j - e_j\right)^2}{e_j} + \sum_{j=1}^{k} \frac{\left(n_j - m_j - e_j^C\right)^2}{e_j^C} \tag{1}$$

where χ^2 is the empirical chi-square value, m_j is the number of units in the j-th population with certain characteristics, e_j is the expected number of units in the j-th population with certain characteristics, n_j is the size of the j-th population, e_j^C is the expected number of units in the j-th population without certain characteristics. The expected number of units in the j-th population with certain characteristics is calculated using the following equation:

$$e_j = n_j \cdot \frac{\sum_{j=1}^{k} m_j}{\sum_{j=1}^{k} n_j} \tag{2}$$

where e_j is the expected number of units in the j-th population with certain characteristics, n_j is the size of the j-th population, m_j is the number of units in the j-th population with certain characteristics.

In the final stage of analysis, simple linear regression analysis will be applied. The estimated simple linear regression model by ordinary least squares method has the following form:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x \tag{3}$$

where $\hat{y}$ is estimated value of the dependent variable, x is independent variable, $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated parameters. In the analysis dependent variable is going to be cumulative proportion of respondents. Cumulative proportion is defined as the sum of respondents with the survey time lower than a certain limit divided by the total number

of respondents. On the other hand, independent variable is the time in which a respondent finished the survey. However, a respondent could finish the survey so that he completes it or break off at the certain point in the survey. Therefore, one regression analysis is conducted where the completion time is independent variable, whereas in the second regression analysis independent variable is the breakoff time. If it were be necessary, the observed variables would be transformed.

3. Analysis of complete responses

3.1. Descriptive statistics of completion times

The observed business web survey covered 1,433 enterprises overall. From that number the survey system registered 780 completed surveys. However, the completion times varied too much with some really strange results. Therefore, further analysis excluded respondents who needed zero (0) seconds to complete the web survey (137 cases) and respondents who needed more than 1,800 seconds (24 cases). In this way, the final number of observed respondents is 619. The distribution of respondents according to their completion time is shown in Table 1.

Table 1. Distribution of respondents according to completion time, all respondents

Time, in min.	Respondents	Percentage	Cumulative percentage
0-5	387	63%	63%
5-10	173	28%	90%
10-15	33	5%	96%
15-20	16	3%	98%
20-30	10	2%	100%
Total	619	100%	-----

Source: Author.

According to the results from Table 1, more than half of respondents needed less than five minutes to complete the survey. Furthermore, 90% of respondents completed it in less than 10 minutes.

Table 2 Distribution of respondents according to completion time, different questionnaire designs regarding presented pictures in the questionnaire

Time, in min.	Positive pictures			Negative pictures			Without pictures		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	125	60%	60%	126	67%	67%	136	62%	62%
5-10	64	30%	90%	47	25%	92%	62	28%	90%
10-15	12	6%	96%	9	5%	96%	12	5%	95%
15-20	7	3%	99%	3	2%	98%	6	3%	98%
20-30	2	1%	100%	4	2%	100%	4	2%	100%
Total	210	100%	-----	189	100%	-----	220	100%	-----

Source: Author.

While Table 1 observes respondents no matter which questionnaire version they are given, in Table 2 respondents' completion times are observed according to the presented pictures in the questionnaire. In this way, respondents who completed questionnaire with positive pictures, with negative pictures or without pictures are observed separately. However, the results are quite similar to the results at the overall level. Therefore, in all three questionnaire versions about 90% of respondents completed their questionnaire version in less than 10 minutes.

Table 3. Distribution of respondents according to completion time, different questionnaire designs regarding the number of questions presented to respondent per questionnaire screen

Time, in min.	One question per screen			Group of questions			All questions		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	151	55%	55%	111	53%	53%	125	95%	95%
5-10	94	34%	88%	73	35%	88%	6	5%	99%
10-15	16	6%	94%	16	8%	95%	1	1%	100%
15-20	8	3%	97%	8	4%	99%	0	0%	100%
20-30	8	3%	100%	2	1%	100%	0	0%	100%
Total	277	100%	-----	210	100%	-----	132	100%	-----

Source: Author.

In Table 3 the distribution of respondents according to the completion time is given again but now the completion times are observed according to the different number of questions presented to a respondent per questionnaire screen. If questionnaire versions with one question per screen and with a group of questions are observed, the completion times are in line with completion times at the overall level. However, in the case when all questions are immediately presented to respondents, almost all respondents (99%) needed less than 10 minutes to complete the questionnaire.

Table 4. Distribution of respondents according to completion time, different questionnaire designs regarding the legal form of enterprises

Time, in min.	Joint stock enterprises			Limited liability enterprises			Simple limited liability enterprises		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	5	42%	42%	359	62%	62%	23	72%	72%
5-10	5	42%	83%	160	28%	90%	8	25%	97%
10-15	2	17%	100%	30	5%	95%	1	3%	100%
15-20	0	0%	100%	16	3%	98%	0	0%	100%
20-30	0	0%	100%	10	2%	100%	0	0%	100%
Total	12	100%	-----	575	100%	-----	32	100%	-----

Source: Author.

Table 4 presents the distribution of respondents according to the completion time when different legal forms of enterprises are taken into account. It has to be emphasized that vast majority of respondents (575 respondents or 93%) were limited liability enterprises and because of that their results are almost the same as at the overall level. In the case of joint stock and simple limited liability enterprises, all respondents managed to complete the questionnaire under 15 minutes time.

Table 5. Distribution of respondents according to completion time, different questionnaire designs regarding the size of enterprises

Time, in min.	Small enterprises			Medium enterprises			Large enterprises		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	373	62%	62%	9	60%	60%	5	71%	71%
5-10	166	28%	90%	5	33%	93%	2	29%	100%
10-15	32	5%	96%	1	7%	100%	0	0%	100%
15-20	16	3%	98%	0	0%	100%	0	0%	100%
20-30	10	2%	100%	0	0%	100%	0	0%	100%
Total	597	100%	-----	15	100%	-----	7	100%	-----

Source: Author.

In Table 5 the size of enterprises is taken into account. Similarly to previous categories, there is also one dominating category here. So, small enterprises have share of 96% in the total number of observed enterprises. All respondents from medium enterprises managed to complete the questionnaire in less than 15 minutes, whereas all respondents from large enterprises managed to complete the questionnaire in less than 10 minutes.

Table 6. Distribution of respondents according to completion time, different questionnaire designs regarding the main activity of enterprises

Time, in min.	Industrial enterprises			Trade enterprises			Service enterprises			Other enterprises		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	109	62%	62%	72	57%	57%	188	65%	65%	18	69%	69%
5-10	48	27%	89%	45	35%	92%	76	26%	91%	4	15%	85%
10-15	11	6%	95%	4	3%	95%	15	5%	96%	3	12%	96%
15-20	5	3%	98%	4	3%	98%	7	2%	99%	0	0%	96%
20-30	3	2%	100%	2	2%	100%	4	1%	100%	1	4%	100%
Total	176	100%	-----	127	100%	-----	290	100%	-----	26	100%	-----

Source: Author.

Table 6 observes respondents' completion time according to the main activity of the enterprises.

It turned out that 92% of respondents from trade enterprises completed the questionnaire in less than 10 minutes. Such respondents from service enterprises were 91%, from industrial enterprises 89% and from other enterprises 85%.

3.2. Comparison of respondents' proportions according to their completion time

In the previous chapter completion time distributions of respondents according to different characteristics have been just reported. In this chapter the differences between completion time distributions of respondents according to different characteristics will be inspected. In order to do that the chi-square test for equality of three or more population proportions is applied. It has been inspected whether the proportions of respondents are equal at the same completion time level across the observed characteristics or not.

Table 7. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding presented pictures in the questionnaire

Time, in min.	Comm. prop.	Exp. res. – positive pictures	Exp. res. – negative pictures	Exp. res. – no pictures	Emp. chi-square	p-value
0-5	0.6252	131.29	118.16	137.54	2.238	0.3267
5-10	0.2795	58.69	52.82	61.49	1.563	0.4577
10-15	0.0533	11.20	10.08	11.73	0.189	0.9098
15-20	0.0258	5.43	4.89	5.69	1.232	0.5401
20-30	0.0162	3.39	3.05	3.55	0.936	0.6262
Total	-----	210	189	220	-----	-----

Source: Author.

In Table 7 the chi-square test for equality of three or more population proportion results between respondents with different presented pictures is shown. The results suggest that at the significance level of 5% the null hypothesis cannot be rejected. This conclusion is valid for all five observed completion time categories. In other words, there is no statistically significant difference in the proportion of respondents in each of the five observed time categories between questionnaires with positive pictures, with negative pictures and without them.

Table 8. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding the number of questions presented to respondent per questionnaire screen

Time, in min.	Comm. prop.	Exp. res. – one question per screen	Exp. res. – group of questions	Exp. res. – all questions	Emp. chi-square	p-value
0-5	0.6252	173.18	131.29	82.53	74.271	<0.0001
5-10	0.2795	77.42	58.69	36.89	45.673	<0.0001
10-15	0.0533	14.77	11.20	7.04	7.758	0.0207
15-20	0.0258	7.16	5.43	3.41	4.855	0.0883
20-30	0.0162	4.47	3.39	2.13	5.571	0.0617
Total	-----	277	210	132	-----	-----

Source: Author.

According to the results given in Table 8, at the significance level of 5%, there is a statistically significant difference in respondents' proportions for different questionnaire designs regarding the number of questions presented to a respondent per questionnaire screen at completion time categories 0–5 minutes, 5–10 minutes and 10–15 minutes. However, it seems that there is no statistically significant difference in respondent's proportions in the last two completion time categories.

Table 9. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding the legal form of enterprises

Time, in min.	Comm. prop.	Exp. res. – joint stock enterprises	Exp. res. – limited liability enterprises	Exp. res. – simple limited liability enterprises	Emp. chi-square	p-value
0-5	0.6252	7.50	359.49	20.01	3.424	0.1805
5-10	0.2795	3.35	160.70	8.94	1.264	0.5316
10-15	0.0533	0.64	30.65	1.71	3.378	0.1847
15-20	0.0258	0.31	14.86	0.83	1.257	0.5334
20-30	0.0162	0.19	9.29	0.52	0.778	0.6778
Total	-----	12	575	32	-----	-----

Source: Author.

Results from Table 9 suggest that there is no statistically significant difference, at the significance level of 5%, in respondent populations according to all observed completion time categories between joint stock enterprises, limited liability enterprises and simple limited liability enterprises.

Table 10. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding the size of enterprises

Time, in min.	Comm. prop.	Exp. res. – small enterprises	Exp. res. – medium enterprises	Exp. res. – large enterprises	Emp. chi-square	p-value
0-5	0.6252	373.25	9.38	4.38	0.278	0.8702
5-10	0.2795	166.85	4.19	1.96	0.223	0.8943
10-15	0.0533	31.83	0.80	0.37	0.448	0.7992
15-20	0.0258	15.43	0.39	0.18	0.605	0.7389
20-30	0.0162	9.64	0.24	0.11	0.375	0.8292
Total	-----	597	15	7	-----	-----

Source: Author.

According to the chi-square test for equality of three or more population proportion results given in Table 10, it can be concluded that at the significance level of 5% the null hypothesis cannot be rejected for all five observed completion time categories. So, there is no difference in respondent distribution according to completion time categories between small, medium and large enterprises.

Table 11. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding the main activity of enterprises

Time, in min.	Comm. prop.	Exp. res. – industrial enterprises	Exp. res. – trade enterprises	Exp. res. – service enterprises	Exp. res. – other enterprises	Emp. chi-square	p-value
0-5	0.6252	110.04	79.40	181.31	16.26	3.025	0.3878
5-10	0.2795	49.19	35.49	81.05	7.27	6.048	0.1093
10-15	0.0533	9.38	6.77	15.46	1.39	3.491	0.3219
15-20	0.0258	4.55	3.28	7.50	0.67	0.930	0.8181
20-30	0.0162	2.84	2.05	4.68	0.42	0.926	0.8192
Total	-----	176	127	290	26	-----	-----

Source: Author.

Finally, in Table 11 the respondents' distributions are compared for industrial, trade, service and other enterprises. It can be concluded that, at the significance level of 5%, there is no statistically significant difference in the respondents' distributions between those four groups of enterprises in all five observed completion time categories.

3.3. Regression modelling of the completion function

In order to estimate the completion function a regression modelling approach is used. In order to keep things straightforward as much as possible, a simple linear regression model is applied. In the regression model the cumulative proportion of respondents is observed as dependent variable, whereas time, given in seconds, is observed as independent variable.

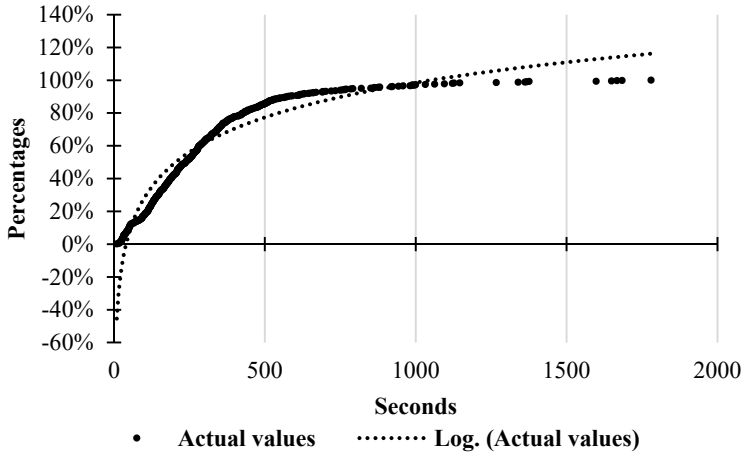

Figure 1. Cumulative distribution of respondents according to completion time and the estimated regression line, all respondents

Source: Author.

According to Figure 1 it is obvious that when all respondents are observed, the relationship between the cumulative proportion of respondents' variable and the time variable is not linear. However, if logarithmic values of the time variable are used, the resulting function is quite similar to the actual distribution of the actual values. In order to be able to compare the results for different characteristics of respondents, in the further analysis only regression models with logarithmic values of the time variable as independent variables are observed.

Still, it has to be emphasized that some limitations appeared due to the data transformation. First of all, by transforming values into logarithmic values the variables completion time and breakoff time are not anymore given in seconds. Therefore, the results are not going to have such intuitive interpretation unless they are transformed back to seconds. Furthermore, the logarithmic transformation could have impact on the data distribution (Feng et al., 2014).

Table 12. Regression analysis results, dependent variable cumulative proportion of respondents, independent variable ln time (in seconds), completion time observed, results for all observed categories of questionnaire designs

Questionnaire designs	No. of res.	No. of unique times	R square	Reg. stand. error	Intercept		Variable ln time	
					Estim.	p-value	Estim.	p-value
Total	619	397	0.9260	0.0807	-1.1253	<0.0001	0.3056	<0.0001
Pictures								
Positive pictures	210	170	0.9081	0.0903	-1.2829	<0.0001	0.3281	<0.0001
Negative pictures	189	162	0.9238	0.0796	-1.1201	<0.0001	0.3062	<0.0001
Without pictures	220	183	0.9350	0.0745	-1.0048	<0.0001	0.2863	<0.0001
Questions								
One question per screen	277	223	0.9327	0.0771	-2.0480	<0.0001	0.4489	<0.0001
Group of questions	210	183	0.9692	0.0508	-2.0379	<0.0001	0.4496	<0.0001
All questions	132	89	0.9139	0.0870	-0.8060	<0.0001	0.3200	<0.0001
Legal form								
Joint stock enterprises	12	11	0.7297	0.1712	-1.0967	0.0094	0.2945	0.0008
Limited liability ent.	575	381	0.9260	0.0808	-1.1110	<0.0001	0.3028	<0.0001
Simple limited liab. ent.	32	31	0.9506	0.0642	-1.4502	<0.0001	0.3757	<0.0001
Size								
Small enterprises	597	389	0.9260	0.0807	-1.1407	<0.0001	0.3078	<0.0001
Medium enterprises	15	15	0.7981	0.1390	-1.0548	0.0004	0.2963	<0.0001
Large enterprises	7	7	0.8855	0.1144	-0.5891	0.0276	0.2623	0.0016
Main activity								
Industrial enterprises	176	154	0.9260	0.0797	-1.0487	<0.0001	0.2910	<0.0001
Trade enterprises	127	116	0.8998	0.0944	-1.1304	<0.0001	0.3032	<0.0001
Service enterprises	290	225	0.9216	0.0831	-1.1628	<0.0001	0.3131	<0.0001
Other enterprises	26	26	0.9614	0.0590	-0.8494	<0.0001	0.2650	<0.0001

Source: Author.

In Table 12 the main regression analysis results for the estimated linear regression models are shown. Each row in Table 12 is related to one linear regression model. According to coefficient determination (R square) values, it can be concluded that all estimated regression models are highly representative. Furthermore, all estimated parameters of the estimated linear regression models are statistically significant at the significance level of 5%.

Table 13. Speeding and completion times estimates based on the regression models, completion time observed, results for all observed categories of questionnaire designs

Questionnaire designs	Speeding time				Completion time			
	Sec.	Min.	No. of faster res.	Per. of faster res.	Sec.	Min.	No. of slower res.	Per. of slower res.
Total	40	0.66	41	6.62%	1,047	17.45	16	2.58%
Pictures								
Positive pictures	50	0.83	13	6.19%	1,052	17.53	5	2.38%
Negative pictures	39	0.65	14	7.41%	1,017	16.94	6	3.17%
Without pictures	33	0.56	17	7.73%	1,099	18.32	4	1.82%
Questions								
One question per screen	96	1.60	4	1.44%	889	14.82	16	5.78%
Group of questions	93	1.55	2	0.95%	861	14.34	13	6.19%
All questions	12	0.21	1	0.76%	283	4.71	7	5.30%
Legal form								
Joint stock enterprises	41	0.69	1	8.33%	1,235	20.58	0	0.00%
Limited liability enterprises	39	0.65	38	6.61%	1,066	17.77	15	2.61%
Simple limited liability enterprises	47	0.79	2	6.25%	680	11.33	1	3.13%
Size								
Small enterprises	41	0.68	40	6.70%	1,047	17.46	16	2.68%
Medium enterprises	35	0.59	1	6.67%	1,027	17.12	0	0.00%
Large enterprises	9	0.16	0	0.00%	427	7.12	0	0.00%
Main activity								
Industrial enterprises	37	0.61	11	6.25%	1,143	19.04	3	1.70%
Trade enterprises	42	0.69	10	7.87%	1,126	18.77	2	1.57%
Service enterprises	41	0.68	18	6.21%	1,001	16.68	9	3.10%
Other enterprises	25	0.41	2	7.69%	1,075	17.91	1	3.85%

Source: Author.

Based on linear regression results given in Table 12, too fast and too slow respondents could be identified. According to estimated linear regression models, too fast respondents are those respondents who completed the survey in less time when the cumulative proportion of respondents in the regression model is equal to 0%. On the other hand, too slow respondents are those respondents who completed the survey in longer time when the cumulative proportion of respondents in the regression model is equal to 100%.

Table 13 presents limit times when a respondent is considered to be too slow or too fast. When all respondents are observed, the limit for speeders is 40 seconds and for too slow respondents 1,047 seconds (about 17 and half minutes). In this way, 41 speeders and 16 too slow respondents are detected who can be omitted from the analysis to get more representative survey results. In Table 13 such results are given for other regression models as well. The results have shown that the most respondents should be omitted at the small enterprises and no respondents at the large enterprises.

4. Analysis of breakoffs

4.1. Descriptive statistics of breakoff times

While in Chapter 4 completion times were observed, in this chapter the breakoff times are the focus. The analysis will be conducted oin analogous way as before. In this way, the direct comparison of the results between those two analyses is possible. Overall 219 respondents started the survey but at some point they broke off.

Table 14. Distribution of respondents according to breakoff time, all respondents

Time, in min.	Breakoffs	Percentage	Cumulative percentage
0-5	203	93%	93%
5-10	11	5%	98%
10-15	3	1%	99%
15-20	0	0%	99%
20-30	2	1%	100%
Total	219	100%	-----

Source: Author.

In Table 14 distribution of all respondents according to the breakoff time is given. As it was expected, the respondents broke off much quicker than the respondents could complete the questionnaire. About 93% of respondents broke off in less than 5 minutes, whereas 98% of respondents broke off in less than 10 minutes.

Table 15. Distribution of respondents according to breakoff time, different questionnaire designs regarding presented pictures in the questionnaire

Time, in min.	Positive pictures			Negative pictures			Without pictures		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	71	95%	95%	68	92%	92%	64	91%	91%
5-10	2	3%	97%	4	5%	97%	5	7%	99%
10-15	1	1%	99%	1	1%	99%	1	1%	100%
15-20	0	0%	99%	0	0%	99%	0	0%	100%
20-30	1	1%	100%	1	1%	100%	0	0%	100%
Total	75	100%	-----	74	100%	-----	70	100%	-----

Source: Author.

In Table 15 distributions of respondents according to their breakoff time for different questionnaire designs regarding presented pictures in the questionnaire are given. The results are quite similar to the results at the overall level.

Table 16. Distribution of respondents according to breakoff time, different questionnaire designs regarding the number of questions presented to respondent per questionnaire screen

Time, in min.	One question per screen			Group of questions			All questions		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	144	94%	94%	57	90%	90%	2	100 %	100%
5-10	8	5%	99%	3	5%	95%	0	0%	100%
10-15	1	1%	99%	2	3%	98%	0	0%	100%
15-20	0	0%	99%	0	0%	98%	0	0%	100%
20-30	1	1%	100%	1	2%	100%	0	0%	100%
Total	154	100%	-----	63	100%	-----	2	100%	-----

Source: Author.

According to the results in Table 16, about 99% of respondents who got questionnaire version with one question per questionnaire screen broke off in less than 10 minutes, whereas 95% of respondents who got questionnaire version with a group of questions per questionnaire screen broke off in the same period. Unfortunately, there were only two breakoffs in the case of the questionnaire when all questions were presented at once.

Table 17. Distribution of respondents according to breakoff time, different questionnaire designs regarding the legal form of enterprises

Time, in min.	Joint stock enterprises			Limited liability enterprises			Simple limited liability enterprises		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	5	83%	83%	188	93%	93%	10	100%	100%
5-10	1	17%	100%	10	5%	98%	0	0%	100%
10-15	0	0%	100%	3	1%	99%	0	0%	100%
15-20	0	0%	100%	0	0%	99%	0	0%	100%
20-30	0	0%	100%	2	1%	100%	0	0%	100%
Total	6	100%	-----	203	100%	-----	10	100%	-----

Source: Author.

Table 17 presents the distribution of respondents according to the breakoff time when different legal forms of enterprises are taken into account. The vast majority of respondents (203 respondents or 93%) were limited liability enterprises. Consequently, their results are the same to the results at the overall level. In the case of joint stock enterprises all respondents have the breakoff time lower than 10 minutes, whereas respondents from simple limited liability enterprises broke off in less than 5 minutes.

Table 18. Distribution of respondents according to breakoff time, different questionnaire designs regarding the size of enterprises

Time, in min.	Small enterprises			Medium enterprises			Large enterprises		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	188	92%	92%	11	100 %	100%	4	100 %	100%
5-10	11	5%	98%	0	0%	100%	0	0%	100%
10-15	3	1%	99%	0	0%	100%	0	0%	100%
15-20	0	0%	99%	0	0%	100%	0	0%	100%
20-30	2	1%	100%	0	0%	100%	0	0%	100%
Total	204	100%	-----	11	100%	-----	4	100%	-----

Source: Author.

According to the results in Table 18, all respondents from medium and large enterprises broke off in less than 5 minutes. About 98% respondents from small enterprises broke off in 10 minutes.

Table 19. Distribution of respondents according to breakoff time, different questionnaire designs regarding the main activity of enterprises

Time, in min.	Industrial enterprises			Trade enterprises			Service enterprises			Other enterprises		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	68	96%	96%	46	94%	94%	80	89%	89%	9	100%	100%
5-10	2	3%	99%	1	2%	96%	8	9%	98%	0	0%	100%
10-15	1	1%	100%	1	2%	98%	1	1%	99%	0	0%	100%
15-20	0	0%	100%	0	0%	98%	0	0%	99%	0	0%	100%
20-30	0	0%	100%	1	2%	100%	1	1%	100%	0	0%	100%
Total	71	100%	-----	49	100%	-----	90	100%	-----	9	100%	-----

Source: Author.

In Table 19 respondents' breakoff times are observed according to the main activity of the enterprises. It has been shown that 99% of respondents from industrial enterprises broke off the questionnaire in less than 10 minutes. Such respondents from service enterprises were 99%, and from trade enterprises 96%. All respondents from other enterprises broke off in less than 5 minutes.

4.2. Comparison of respondents' proportions according to their breakoff time

In this chapter the distributions of respondents according to their breakoff time levels are compared by the chi-square test for equality of three or more population proportions. Unfortunately, due to lack of data it was not possible to conduct the chi-square test for 15–20 minutes time level.

Table 20. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding presented pictures in the questionnaire

Time, in min.	Comm. prop.	Exp. break. – positive pictures	Exp. break. – negative pictures	Exp. break. – no pictures	Emp. chi-square	p-value
0-5	0.9269	69.52	68.59	64.89	0.667	0.7165
5-10	0.0502	3.77	3.72	3.52	1.555	0.4596
10-15	0.0137	1.03	1.01	0.96	0.003	0.9986
15-20	-----	-----	-----	-----	-----	-----
20-30	0.0091	0.68	0.68	0.64	0.948	0.6224
Total	-----	75	74	70	-----	-----

Source: Author.

In Table 20 the chi-square tests for equality of three or more population proportion results where breakoff time is observed, for different questionnaire designs regarding presented pictures in the questionnaire are observed. At the significance level of 5%, the null hypothesis cannot be rejected at all observed time levels. So, the structure of breakoff is the same at all three questionnaire designs across all given time levels.

Table 21. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding the number of questions presented to respondent per questionnaire screen

Time, in min.	Comm. prop.	Exp. break. – one question per screen	Exp. break. – group of questions	Exp. break. – all questions	Emp. chi-square	p-value
0-5	0.9269	142.75	58.40	1.85	0.765	0.6820
5-10	0.0502	7.74	3.16	0.10	0.124	0.9397
10-15	0.0137	2.11	0.86	0.03	2.138	0.3433
15-20	-----	-----	-----	-----	-----	-----
20-30	0.0091	1.41	0.58	0.02	0.453	0.7972
Total	-----	154	63	2	-----	-----

Source: Author.

According to Table 21, at the significance level of 5%, there is no statistically significant difference in the proportion of respondents' breakoff at all observed time for different questionnaire designs regarding the number of questions presented to a respondent per a questionnaire screen.

Table 22. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding the legal form of enterprises

Time, in min.	Comm. prop.	Exp. break. – joint stock enterprises	Exp. break. – limited liability enterprises	Exp. break. - simple limited liability enterprises	Emp. chi-square	p-value
0-5	0.9269	5.56	188.17	9.27	1.567	0.4569
5-10	0.0502	0.30	10.20	0.50	2.238	0.3266
10-15	0.0137	0.08	2.78	0.14	0.240	0.8870
15-20	-----	-----	-----	-----	-----	-----
20-30	0.0091	0.05	1.85	0.09	0.159	0.9235
Total	-----	6	203	10	-----	-----

Source: Author.

Table 22 reveals that there is no statistically significant difference in the proportion of respondents' breakoffs according to the observed time categories between respondents from enterprises with a different legal form.

Table 23. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding the size of enterprises

Time, in min.	Comm. prop.	Exp. break. – small enterprises	Exp. break. – medium enterprises	Exp. break. – large enterprises	Emp. chi-square	p-value
0-5	0.9269	189.10	10.20	3.71	1.269	0.5301
5-10	0.0502	10.25	0.55	0.20	0.852	0.6532
10-15	0.0137	2.79	0.15	0.05	0.224	0.8942
15-20	-----	-----	-----	-----	-----	-----
20-30	0.0091	1.86	0.10	0.04	0.148	0.9285
Total	-----	204	11	4	-----	-----

Source: Author.

Results from Table 23 are in line with all the previous results. Again, it can be concluded that there is no statistically significant difference in respondent's proportions according to the observed time levels. So, this conclusion is valid for enterprises of different size as well.

Table 24. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding the main activity of enterprises

Time, in min.	Comm. prop.	Exp. break. – industrial enterprises	Exp. break. – trade enterprises	Exp. break. – service enterprises	Exp. break. – other enterprises	Emp. chi-square	p-value
0-5	0.9269	65.81	45.42	83.42	8.34	3.730	0.2921
5-10	0.0502	3.57	2.46	4.52	0.45	4.933	0.1767
10-15	0.0137	0.97	0.67	1.23	0.12	0.334	0.9536
15-20	-----	-----	-----	-----	-----	-----	-----
20-30	0.0091	0.65	0.45	0.82	0.08	1.465	0.6904
Total	-----	71	49	90	9	-----	-----

Source: Author.

Finally, in Table 24 distributions of respondents' breakoffs are compared for enterprises of different main activity. The chi-square results lead to the conclusion that there is no statistically significant difference in respondents' proportions at all observed breakoff time levels.

4.3. Regression modelling of the breakoff function

In order to estimate the breakoff function linear regression modelling is applied. According to Figure 2, where cumulative distribution of all respondents according to breakoff time is shown, the relation between these two variables is not linear. Because of that, in the linear regression model cumulative proportion of respondents is observed as dependent variable, whereas the logarithm of time is observed as independent variable.

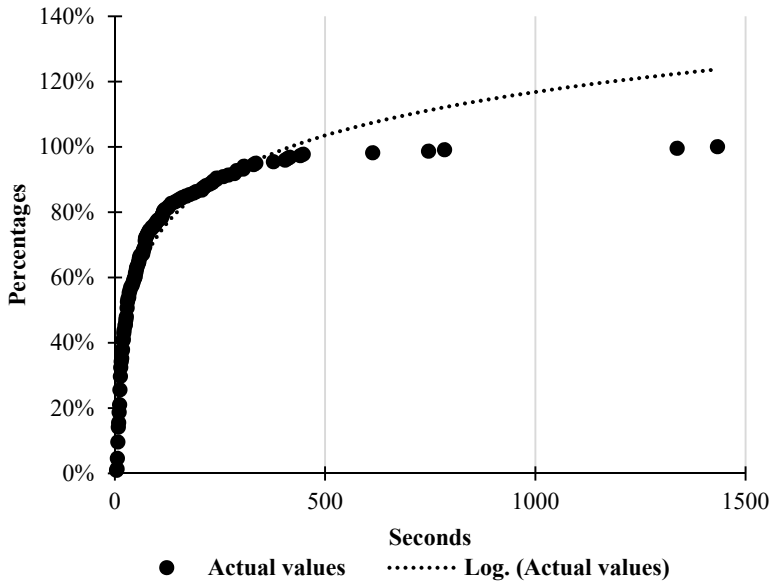

Figure 2. Cumulative distribution of respondents according to breakoff time, all respondents
Source: Author.

When actual and regression are compared in Figure 2, it can be concluded that the regression function follows the actual values quite well. The results in Table 25 confirmed that the estimated regression model fits well the actual values.

Table 25. Regression analysis results, dependent variable cumulative proportion of respondents, independent variable ln time (in seconds), breakoff time observed, results for all observed categories of questionnaire designs

Questionnaire designs	No. of. break.	No. of unique times	R square	Reg. stand. error	Intercept		Variable ln time	
					Estim.	p-value	Estim.	p-value
Total	219	110	0.9467	0.0568	-0.1597	<0.0001	0.1923	<0.0001
Pictures								
Positive pictures	75	54	0.9409	0.0671	-0.2235	<0.0001	0.2095	<0.0001
Negative pictures	74	53	0.9285	0.0728	-0.1701	<0.0001	0.1989	<0.0001
Without pictures	70	50	0.9829	0.0358	-0.2421	<0.0001	0.2050	<0.0001
Questions								
One ques. per screen	154	74	0.9061	0.0794	-0.1106	0.0008	0.1941	<0.0001
Group of questions	63	56	0.9727	0.0464	-0.3839	<0.0001	0.2139	<0.0001
All questions	2	2	1.0000	0.0000	-1.0019	-----	0.4666	-----
Legal form								
Joint stock ent.	6	6	0.9779	0.0519	-0.3996	0.0065	0.2280	0.0002
Limited liability ent.	203	104	0.9434	0.0585	-0.1466	<0.0001	0.1900	<0.0001
Simple lim. liab. ent.	10	10	0.8510	0.1239	-0.4455	0.0192	0.2786	0.0001
Size								
Small enterprises	204	104	0.9425	0.0598	-0.1524	<0.0001	0.1905	<0.0001
Medium enterprises	11	11	0.9810	0.0438	-0.3273	<0.0001	0.2523	<0.0001
Large enterprises	4	4	0.9429	0.0945	-0.2907	0.2225	0.2201	0.0290
Main activity								
Industrial enterprises	71	50	0.9752	0.0432	-0.2922	<0.0001	0.2217	<0.0001
Trade enterprises	49	37	0.9256	0.0781	-0.1828	<0.0001	0.2008	<0.0001
Service enterprises	90	56	0.9564	0.0587	-0.1731	<0.0001	0.1930	<0.0001
Other enterprises	9	9	0.9006	0.1026	-0.4171	0.0133	0.2773	0.0001

Source: Author.

In Table 25, except for all respondents, main regression results for different categories of respondents are given as well. Almost all estimated regression model estimated regression coefficients are statistically significant at the significance level. Only in the regression model where large enterprises are observed, the constant term in not statistically significant at 5%. However, this is mainly due to fact that the regression model was estimated based on only four respondents.

Table 26. Speeding and completion times estimates based on the regression models, breakoff time observed, results for all observed categories of questionnaire designs

Questionnaire designs	Speeding breakoff time				Completion breakoff time			
	Sec.	Min.	No. of faster break.	Per. of faster break.	Sec.	Min.	No. of slower break.	Per. of slower break.
Total	2	0.04	0	0.00%	417	6.94	7	3.20%
Pictures								
Positive pictures	3	0.05	0	0.00%	343	5.72	4	5.33%
Negative pictures	2	0.04	0	0.00%	358	5.97	4	5.41%
Without pictures	3	0.05	0	0.00%	428	7.13	1	1.43%
Questions								
One question per screen	2	0.03	0	0.00%	306	5.09	9	5.84%
Group of questions	6	0.10	3	4.76%	644	10.74	3	4.76%
All questions	9	0.14	0	0.00%	73	1.22	0	0.00%
Legal form								
Joint stock enterprises	6	0.10	0	0.00%	464	7.73	0	0.00%
Limited liability enterprises	2	0.04	0	0.00%	418	6.97	7	3.45%
Simple limited liability enterprises	5	0.08	0	0.00%	179	2.99	1	10.00%
Size								
Small enterprises	2	0.04	0	0.00%	423	7.05	7	3.43%
Medium enterprises	4	0.06	0	0.00%	193	3.21	0	0.00%
Large enterprises	4	0.06	0	0.00%	352	5.86	0	0.00%
Main activity								
Industrial enterprises	4	0.06	0	0.00%	340	5.67	2	2.82%
Trade enterprises	2	0.04	0	0.00%	362	6.03	3	6.12%
Service enterprises	2	0.04	0	0.00%	436	7.26	2	2.22%
Other enterprises	5	0.08	0	0.00%	166	2.76	1	11.11%

Source: Author.

In the same way as at the estimated completion functions, estimated breakoff functions can be used to identify respondents who break off too quickly or too slowly. However, according to Table 26, only respondents who had the questionnaire version where the group of questions were presented to a respondent per questionnaire screen had speeders. On the other hand, some too slow breakoff respondents were identified at some questionnaire designs.

5. Discussion

In the paper the completion and breakoff functions of a business web survey are estimated. The estimations were conducted at the overall level, for all respondents together, but also according to certain characteristics of respondents. The estimates for the completion functions can be found in Table 12, whereas the estimates for the breakoff functions are given in Table 25. Due to paper length limits estimated functions are graphically presented only at the overall level for all respondents. In this way, in Figure 1 the completion function for all respondents is given, whereas in Figure 2 the breakoff function is shown.

According to the first research hypothesis, the optimal survey lengths, estimated by the completion and breakoff functions, are different for different questionnaire designs and for respondents of different characteristics. The optimal survey length here is defined as the time in which all respondents should have completed the survey. In this way, the optimal survey length is observed as the longest time in which respondents should have completed the survey. The longest time can be calculated by observing the estimated completion functions and calculate times when 100% as value of dependent variable is given.

Optimal survey lengths are given in Table 13. According to the results, the optimal survey length, when all respondents are taken into account, is 1,047 seconds or 17.45 minutes. If only results in Table 13 are observed, it can be concluded that for some questionnaire designs and for some characteristics of respondents the optimal survey length is not so different, whereas in some cases the difference is considerable. The difference in the optimal survey length seems to be rather small in the case of questionnaire designs where different pictures were presented to the respondents and in the case when enterprises are observed according to their main activity. In other cases, there is always one category at which the optimal survey length is considerably lower than in other categories. Because of that, researches have to take into account different questionnaire designs and characteristics of respondents during the process of developing the questionnaire design. In support of this process, the results from Chapter 3.2., where distributions of respondents according to the completion times are observed, should be consulted as well. In this way, generally speaking, the first research hypothesis can be accepted.

The estimated completion and breakoff functions could be used to detect respondents who were too fast or too slow. Too fast respondents did not have enough time for full cognitive perception of questions. Also, respondents could break off before they read the introductory survey page. On the other hand, too slow respondents probably just opened the survey and worked something else. This could happen in business surroundings very often because employees who are participating in the

survey could be interrupted, for example, by e-mail, phone call or other work colleagues. It could happen that a respondent left the survey open and then quickly provides answers, also. In both cases, in cases of too fast and too slow respondents, such respondents should be omitted from the analysis because their answer cannot be declared valid. If such answers would not been omitted, they could have certain impact on survey results and, consequently, wrong conclusions could be made.

In order to detect respondents who were too fast or too slow, the completion and breakoff functions should be used to calculate the limit times. The limit times are calculated by taking into account that the cumulative percentage of respondents is 0% and 100% respectively. In this way, two times are obtained in the completion and breakoff functions. If respondent's survey time is lower than the lower limit, the conclusion is that this respondent was too fast. On the other hand, if the respondents' survey time was longer than the upper limit, the conclusion is that those respondents were too slow. Those limits calculated based on the completion and breakoff functions are presented in Table 13 and Table 26, respectively. Except to omit too fast and too slow respondents which answers could have significant impact on the survey results, the results of this analysis could be used for further improvement of the survey questionnaire design. Finally, it can be concluded that the second research hypothesis can be accepted.

6. Conclusion

The response rates in web surveys tend to be very low. Because of that, researchers should invest more effort to reach some appropriate response rate levels. One of the ways to increase response rates is to carefully design the questionnaire and its length. However, the question is how to know whether the questionnaire is too long or too complicated for respondents or not.

In the paper the completion and breakoff functions are proposed to be used to determine the optimal survey length. Those functions are estimated by observing cumulative proportion distribution of respondents according to their completion and breakoff times in a web business survey conducted on a sample of enterprises in Croatia. In order to keep things as simple as possible and therefore easily interpretable at the same time, a simple linear regression approach to the completion and breakoff functions estimation was used.

After the completion and breakoff functions have been estimated, the possibilities of their use are shown. It has been illustrated how the completion function can be used to estimate the optimal survey length or to estimate time in which respondents, even those who are inexperienced or unfamiliar with the survey, should complete the survey. In the paper additional possible use of the completion and breakoff functions is

presented as well. So, the use of the completion and breakoff functions in detecting too fast and too slow respondents is recommended as well.

However, while the regression diagnostic results have shown that the estimated completion and breakoff functions are of good fit, the estimates are valid only for the observed survey. So, the main limitation of the proposed approach is that a pilot study is needed to be able to estimate the completion and breakoff functions. In further research a way of estimating standardized completion and breakoff functions which could be used in business web survey should be found.

References

1ka, (2020a). Osnovna priporočila: Kaj pomeni kompleksnost ankete? (in Slovenian). Available at: https://www.1ka.si/d/sl/spletne-ankete/osnovna-priporocila/kaj-pomeni-kompleksnost-ankete [02 June 2020].

1ka, (2020b). Osnovna priporočila: Kako dolga naj bo moja anketa? (in Slovenian). Available at: https://www.1ka.si/d/sl/spletne-ankete/osnovna-priporocila/kako-dolga-naj-bo-moja-anketa [02 June 2020].

Blumenberg, C., Zugna, D., Popovic, M., Pizzi, C., Barros, A.J.D., Richiardi, L., (2018). Questionnaire Breakoff and Item Nonresponse in Web-Based Questionnaires: Multilevel Analysis of Person-Level and Item Design Factors in a Birth Cohort, Journal of Medical Internet Research, Vol. 20, No. 12., pp. 1–11, DOI: 10.2196/11046.

Crawford, S. D., Couper, M. P., Lamias, M. J., (2001). Web surveys: Perceptions of burden, Social Science Computer Review, Vol. 19, No. 2, pp. 146–162.

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., Tu, X. M., (2014). Log-transformation and its implications for data analysis, Shanghai archives of psychiatry, Vol. 26, No. 2, pp. 105–109.

Galesic, M., (2006). Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey, Journal of Official Statistics, Vol. 22, No. 2, pp. 313–328.

Groves, R. M., (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys, Public Opinion Quarterly, Vol. 70, No. 5, pp. 646–675.

Göritz, A. S., (2006). Incentives in Web Studies: Methodological Issues and a Review, International Journal of Internet Science, Vol. 1, No. 1, pp. 58–70.

Kish, L., (1995). Survey Sampling, New York: John Wiley & Sons.

Mittereder, F. K., (2019). Predicting and Preventing Breakoff in Web Surveys, dissertation, University of Michigan, pp. 1–762, available at: https://isr.umich.edu/wp-content/uploads/2019/04/Dissertation_FinalSubmit_V03.pdf.

Narodne novine, (2007). The National Classification of Economic Activities. Official Gazette, 16(58).

Narodne novine, (2011). The Enterprises Act. Official Gazette, 20(152).

Narodne novine, (2015). The Accounting Act. Official Gazette, 24(78).

Peytchev, A., (2009). Survey Breakoff, Public Opinion Quarterly, Vol. 73, No. 1, pp. 74–97.

Peytchev, A., (2011). Breakoff and Unit Nonresponse Across Web Surveys, Journal of Official Statistics, Vol. 27, No. 1, pp. 33–47.

Žmuk, B., (2017). Impact of Questionnaire Length and Complexity on Survey Time: Comparison of Two Business Web Questionnaire Versions, Metodološki zvezki - Advances in Methodology and Statistics, Vol. 14, No. 2, pp. 49–74.