What Took You so Long? The Role of Experience as a Determinant of Interview Duration

Kristina Altmann[1] and Tobias Schmidt[2]
[1] Deutsche Bundesbank, Statistics Department , Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany
[2] Deutsche Bundesbank, Research Centre, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany, Tobias.Schmidt@bundesbank.de

Abstract
The focus of this paper is on how the experience of interviewers and respondents influences interview duration. In contrast to existing studies we distinguish experience along several dimensions: 1) across waves of a panel study, 2) across interviews within one wave and 3) across different studies at the same survey institute. Our paper is one of the first to look at multiple measures of experience in a panel context. We analyse the time (per question) it takes respondents in the German Wealth Survey "Panel on Household Finances – PHF" to complete interviews. The survey is an interviewer mediated CAPI interview, which consists of a household interview and a personal interview of every household member at least 16 year old. In 2014, 4,461 households took part in the survey and about half of the survey sample is a panel interviewed for the second time. In our study we find similar patterns to those typically observed in the literature with respect to within study and general interviewer experience. A novel finding is that panel respondents tend to have longer per item interview duration than respondents interviewed for the first time. We do not find clear evidence that across waves panel interviewers tend to have different interview durations per item compared to other interviewers.

Key Words: survey methodology, interview duration, interviewer experience, household surveys

Disclaimer: This paper represent the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

1. Introduction

The time it takes respondents to answer surveys or survey questions is an important topic in survey methodology research. Response time is used to predict data quality, as a proxy for the motivation and commitment of respondents to a survey, or a tool to identify problematic questions. Survey agencies sometimes pay their interviewers based on response time. Because of this there is a lot of research on the determinants of interview duration. Interviewer experience has been identified as a relevant predictor of interview duration. In a seminal contribution Olson and Peytchev (2007) show that interviewers speed up over the course of one study and that more experienced interviewers have shorter interview times. Yan and Tourangeau (2008) find a similar effect for respondents to web surveys, response times per item fall as the interview progresses and respondents' general experience with web surveys decreases response times.

Even though Olson and Peytchev (2007) mention that "… experience over multiple waves of a longitudinal survey" (p. 274) can also play a role, this aspect of experience has not received a lot of attention. To our knowledge we are the first to investigate how panel experience of both respondents and interviewers affects response times. Other aspects of interviewer experience are also considered in our study, in particular previous work experience and the number of interviews in the same wave of the survey prior to the current interview.

We find the patterns typically observed in the literature with respect to within study and general interviewer experience at the survey company. A novel finding, obtained using multi-level models, is that across waves panel interviewers do not tend to have shorter interview times per item than other interviewers. And even more surprising, panel respondents tend to have longer interview times per item than respondents interviewed for the first time.

Our study shows how useful para-data can be (Couper and Kreuter, 2013) for the analysis of interview duration and survey processes more general. The findings indicate that when analysing response times from face-to-face interviews it is important to take the experience and sample composition of the interviewers and respondents into account.

Our paper documents significant differences in interview duration per item for different types of interviewers and households, but remains silent about the specific processes behind the heterogeneous effects. We also leave an assessment of what the different interview times mean, e.g. in terms of data quality, for future research.

In the next section we review the literature on interview duration and experience, before we present the data set and variables for our analysis. In section 4 we report both descriptive results and results from a multi-level model. Section 5 conculdes.

2. Experience and Interview Duration in Face-to-Face Surveys – A Short Review of the Literature

The literature on interview duration is vast, not least because interview duration has been used as a proxy for data quality and response burden (Bassili, 1996; Yan and Tourangeau, 2008; Draisma and Dijkstra, 2004; Loosveldt and Beullens, 2013). Many researchers have analysed the effects of respondent, interviewer and questionnaire characteristics on interview duration and interview duration per item (Couper and Kreuter, 2013; Gummer and Roßmann, 2015; Loosveldt and Beullens, 2013; Heerwegh and Loosveldt, 2002; Yan et al., 2015).

We will focus on one specific aspect of interviewer and respondent characteristics, ie. experience. Studies on how experience and interview duration are linked are not as prevalent as studies on other determinants of interview duration.

Olson and Peytchev (2007) analyse data from three waves of the National Election Studies in the US. They are mainly concerned with how interviewer effects change over the field work period. They find that interviewers speed up as the conduct more interviews within on study wave. Olson and Peytchev (2007) also show that more experienced interviewers have shorter interview times. In their literature review they touch upon some of the reasons why experienced interviewers may have shorter interview times. [1] They cite, among others, Fowler (1991) and Pickery and Loosveldt (2001) to motivate the statement, that interviewers may complete surveys quicker as they become more and more careless in administrating the questionnaire. On the other hand, interviewers may also learn how to administer the survey instrument "[…] and change their behaviours accordingly (However, Cannell, Marquise and Laurent, 1977)" (Olson and Peytchev, 2007, p. 274).[2]

Wuyts et al. (2018) analyse data from the Belgium ESS and show, that more experienced interviewers have shorter interviews. They measure interview duration as the average number of questions per minute and interviewer experience as experience working as an interviewer. They also find that within one study interviewers get quicker as they conduct more interviews.

Olson and Bilgen (2011) analyse the link between interview experience and acquiescence behaviour. Their descriptive statistics for the 2000 American National Election Survey (ANES) and the 2006 General Social Survey (GSS) show that experienced interviewers are on average significantly quicker than inexperienced interviewers. The difference between the two groups is 6 minutes in the ANES and 8 minutes in the GSS. Loosveldt and Beullens (2013) find that interviewers speed up as they conduct more interviews within one wave of the European Social Survey. They also document that the scope of interviewer induced variance in interview duration varies a lot across countries, indicating that the results for one country may not hold in other settings.

A contribution of our study is to look at the relationship between the panel experience of interviewers and interview duration. There is some evidence that prior experience of interviewers within the same study has an effect on data quality, [3] but to our knowledge nobody has thus far linked interviewer experience across waves within the same study ("panel interviewer") to interview duration. We thus now turn to the literature on respondents' experience and interview duration.

Yan and Tourangeau (2008) look at the experience of respondents instead of interviewers in a self-administered web-survey. They find an effect for respondents to web surveys similar to that for interviewers in face-to-face surveys: response times per item fall as the interview progresses and respondents' general experience with web surveys decreases response times.

[1] They measure interviewer experience as working as an interviewer prior to the survey. Since the survey they are analysing is not a panel, they cannot look at how experience across waves of the same survey influence response times. Their measure of interview duration is the total length of the interviews in minutes.

[2] We do not include a review of the literature on the mechanisms through which experience influences interview duration, as this aspect is beyond the scope of our analysis. However, the papers cited above typically also provide some assessment of the mechanisms (acquiescence, satisficing, interview fatigue, learning…) behind the differential interview duration.

[3] Bailar et al. (1997) analyse the experience with the March supplement of the CPS and find that interviewers participating in several March supplement interviews have higher item non-response rates on income questions and have a higher share of respondents with only one type of income than less experienced interviewers. If the behaviour related to those outcomes can speed up the interview, the interviews may be shorter for experienced interviewers.

Toepoel et al. (2008) use two Dutch panel surveys and document that "trained respondents", ie. those with prior experience within the panel survey, are quicker in completing web-survey questions. They argue that this may be due to "satisficing behaviour" and show that experienced respondents pick the first item of questions with several response options more frequently than less experienced respondents. Gummer and Roßmann (2015) confirm that less experienced respondents take more time to answer questions than experienced respondents in a pooled data-set of 21 web surveys from Germany.
The studies cited above rely mainly on total interview duration or interview duration for specific modules of a larger study. Couper and Kreuter (2013) look at key stroke data for specific questions in the National Survey of Family Growth. They analyse the determinants of item-level response times using characteristics of the item, respondent and interviewer. Prior experience of interviewers with Computer Assisted Interviewing (CAI) seems to matter. Interviewers with this type of experience are quicker in administrating the questionnaire to female respondents (statistically significant) and males (same sign, but not statistically significant).
In summary, the literature clearly shows that both interviewers' and respondents' experience matters for interview duration.

3. Data and Descriptive Statistics

For our analysis we use data from the German wealth survey "Panel on Household Finances (PHF)" [4] and in particular the rich para-data collected during the survey process. The face-to-face CAPI survey is set up as a panel study and has been conducted in 2010 and 2014. About half of the 4,161 households interviewed in the 2014 wave, are panel households participating for the second time. Both surveys have been conducted by the same survey agency allowing us to trace not only panel households but also the panel status of the interviewers. The para-data includes information on interviewer characteristics, but most importantly also measures of interview duration. Keystroke data is unfortunately not available, but we can calculate interview duration per item from the build in CAPI timer and the obtained micro-data and in doing so control for the number of questions asked in the heavily filtered survey instrument.
The PHF survey has several different parts, one long part related to a household's wealth, liabilities and income, and one part referring to each household member's employment status, income and pensions. Every person at least 16 years old is supposed to answer the short individual level questionnaire, while the household part is answered by only one person, i.e. the person who knows best about the finances of the given household. This so-called "financially knowledgeable person (FKP)" is determined at the beginning of the interview with a structured questionnaire. This very first part of the interview is not timed. For the FKP the short personal-level interview is combined with the long interview referring to the household-level variables to form one full interview. For personal level interviews it is possible to conduct the interview by telephone and proxy interviews are possible. The full interview with the FKP, which we use for our analysis, is required to be conducted in person and with the selected FKP.

[4] For details on the survey see www.bundesbank.de/phf-research

3.1 Measuring Interview Duration

Our main variable of interest is the duration of the interview with the FKP. The duration is determined with a latent timer in the CAPI, which starts once the interviewer calls up the "Greetings and data protection message" at the beginning of the interview and ends with the closing screen in the CAPI. The interviewers' salary per interview is to a large degree based on the duration of the interview in minutes, which makes us confident that the timing data is of high quality. The payment scheme of course also provides an incentive for the interviewer not to rush respondents through the survey.

Because the questionnaire uses a lot of filtering to spare respondents to answer questions that are not relevant for them, the number of questions answered per respondent varies a lot. We take this into account and will not only analyse the interview duration as such, but the interview duration per question/item.[5]

The shortest interview in terms of the number of questions to be answered had 124 questions and the longest 460 questions. In most parts of the survey there is little filtering based on the panel membership of the household/individual. An exception is the employment section, where individuals retired in both waves are not again asked about their previous employment status in wave 2. This is certainly one of the reasons why the mean number of questions answered by panel respondents is at 209 questions significantly lower than the mean for the new respondents, which stands at 256 questions.

We follow Mayerl (2013) and clean the raw timing data by dropping cases with values outside a range of plus/minus two standard deviations from the mean. For the total interview time 184 out of 4,461 observations are outside this range and an additional 184 observations for the interview duration per item. After this cleaning procedure we have 4,093 interviews with FKPs left in the analysis sample.

After applying these corrections we get the following distributions (see figures 1 and 2) for total interview duration per interview with a financially knowledgeable person and interview duration per-item. Both distributions are close to being normally distributed. The average interview duration is 65.3 minutes, with a standard deviation of 18.6 minutes. The median is at 63.7 minutes. For the interview duration per item (figure 2) the numbers are 17.0 sec (mean), 4.6 sec (sd) and 16.5 sec (median).

[5] Throughout the analysis all questions types and information screens will count as one item.

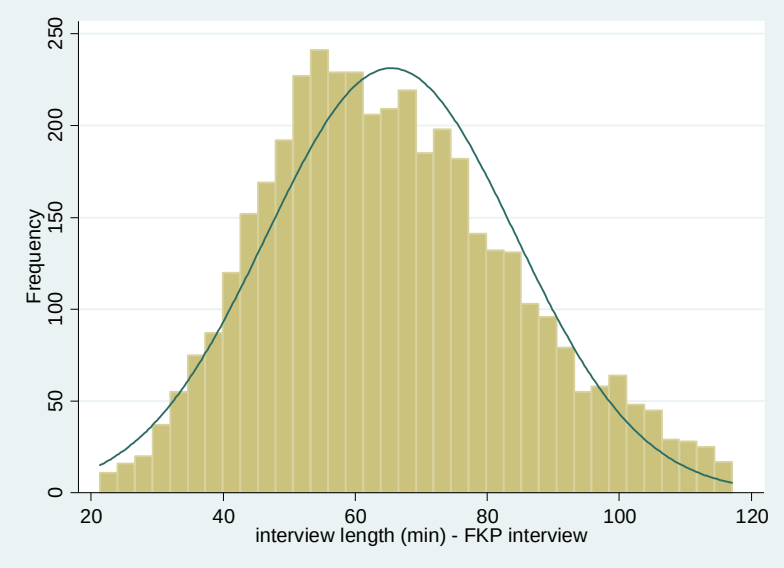


Figure 1: Interview Duration per Interview with a Financially Knowledgeable Person

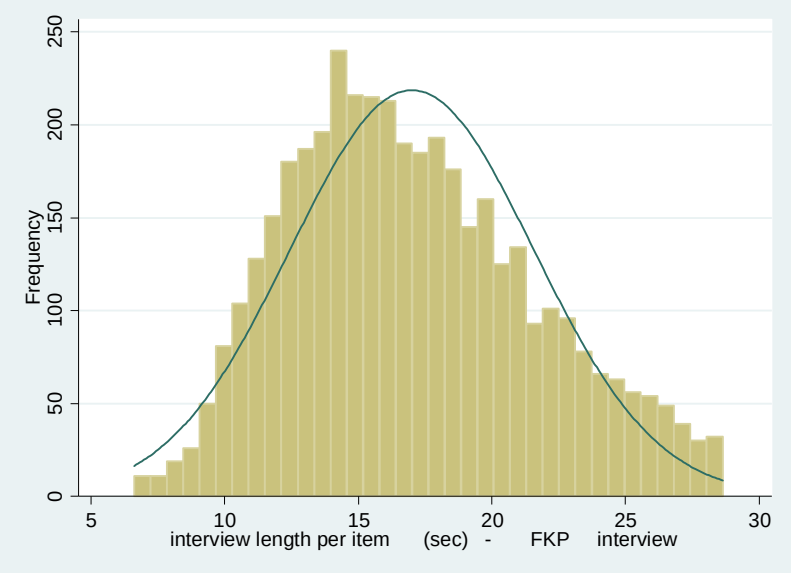


Figure 2: Interview Duration per Item in sec with Financially Knowledgeable Person

3.2 Measuring the Experience of Respondents and Interviewers

Our main interest in this paper is on how experience influences interview duration. We look at this issue with different indicators of experience:

a) Experience of the interviewers within the same study, but across waves
The data allow us to differentiate between interviewers that have not worked on the German wealth survey project in previous waves ("new interviewers") and those who have conducted interviews in previous waves ("panel interviewers").
About 30% of the 244 interviewers that conducted at least one interview in wave two, are panel interviewers, 70% are working on the German wealth survey for the first time in wave two.

b) Experience of the interviewers with regards to years worked as interviewer
Unfortunately we cannot observe the number of years a person has been working as an interviewer. Since we only have data from one survey company, we only know how long the interviewer has been working for that particular company. We are confident that the measure we have is a good proxy for total number of years, though. [6]
The average number of years worked as an interviewer is 4.3, for the 244 interviewers that conducted at least one interview in wave two.

c) Experience of the interviewers within the same study and survey wave
This measure of experience is simply the number of the interview for one interviewer within the same wave of the German wealth survey. The specific sequence of the interviews determines this level of experience: the first interview of each interviewer gets a one, the second and two, etc.
On average the interviewers conducted 15.6 interviews each. However, the heterogeneity of interviewers in that respect is very large. One percent of all interviewers conducted more than 75 interviews, but also 25% at most 5; 236 interviews have less than 50 interviews.

d) Experience of the respondent within the same study across waves
Similar to the measure for interviewers, we construct a dummy variable taking the value one if the respondent was interviewed in previous survey waves ("panel household") [7] or not ("refresher household").
The split is almost in the middle, with 47% of households in our final sample from the panel sample and 53% being refresher households.

The goal of the survey company was to match panel households with panel interviewers. Due to the long gap between the two survey waves (three years), this was not possible in all cases. In the end the following structure emerged in our sample:

[6] An even better measure of overall interviewer experience would be the number of interviews conducted or surveys worked on, over all the years of working as an interviewer. In Germany, most interviewers are freelance and work for several survey companies at the same time. Compiling these types of indicators would thus be even more challenging than the collection of the indicators we are using.

[7] We will talk about "refresher households" and "panel households", because we analyse the household interview that includes the personal interview of the financially knowledgeable person (FKP), but not the interviews of other persons.

Table 1: Match between (Panel) Interviewers and (Panel) Respondents

	New Interviewer	Panel Interviewer	Total
Refresher Household	36 %	17 %	53 %
Panel Household	15 %	32 %	47 %
Total	51%	49%	100%

About one third of the possible interviewer-respondent combinations are panel-panel, one third non-panel-non-panel and one third panel- non-panel combinations.

3.3 Control Variables

In the multi-variate analysis presented in the second part of the results section we control for socio-demographic characteristics like gender and age of the respondent and interviewer, respondents' education, household size, and an estimate of household total wealth. We also make use of para-data collected by the interviewers on how suspicious and interested respondents were before the interview, whether respondents had difficulties answering or problems to express themselves and whether documents were used during the interview. Slightly more than one third of the interviewers is female (34%), the average age of interviewers is 58 years and 83% are older than 50 years. The financially knowledge persons are on average only slightly younger at 55 years. The share of female FKPs is at 43% higher than the share of female interviewers. The education of the FKPs is medium (42%) to high (41%). The unweighted self-assessed wealth of the households sums to an average of 288,400 euros (median 100,000). Most households are two person households (45%), only 15% of households have 4 or more members.
One fifth of households reported to be suspicious before the interview, 41% reported to be interested in the survey and 23% reported to be very interested. The survey instrument seems to be well suited for the respondents, only 6% reported difficulties with answering the questions, 37% had no difficulties at all and 45% small difficulties. The interviewers also rated the ability of respondents to express themselves as high (38%) or very high (55%). Documents were used by about 38% of households, 6% used them often throughout the interview.

4. Results

In this section we will review the main results of our analysis. We will first look at the descriptive evidence looking at individual types of experience before moving on to a multi-level analysis that takes several aspects of experience into account at the same time. All the results presented here refer to the interview duration per item.

4.1 Descriptive Analysis

The descriptive analysis will address the five different types of experience introduced above, starting with interviewer experience and then moving to respondent experience. [8]

[8] Some descriptive statistics for the interview duration by socio-demographic characteristics of respondents and interviewers are presented in the appendix.

4.1.1 Interviewers' Experience

The main novelty of our paper is to introduce a type of experience so far neglected in the literature on interview duration, the panel experience of the interviewers. This indicator alone cannot explain differences in interview duration per item across households. The mean and median durations per item are almost the same for panel interviewers and interviewers new to the PHF survey.

Table 2: Average Interview Duration per Item, by Interviewers' Experience with PHF Study

	Number of Households	Mean	Median	Variance
Panel interviewers	1,990	17.05	16.48	22.37
Interviewers new to PHF	2,103	16.97	16.49	19.49
All Interviewers	4,093	17.01	16.49	20.88

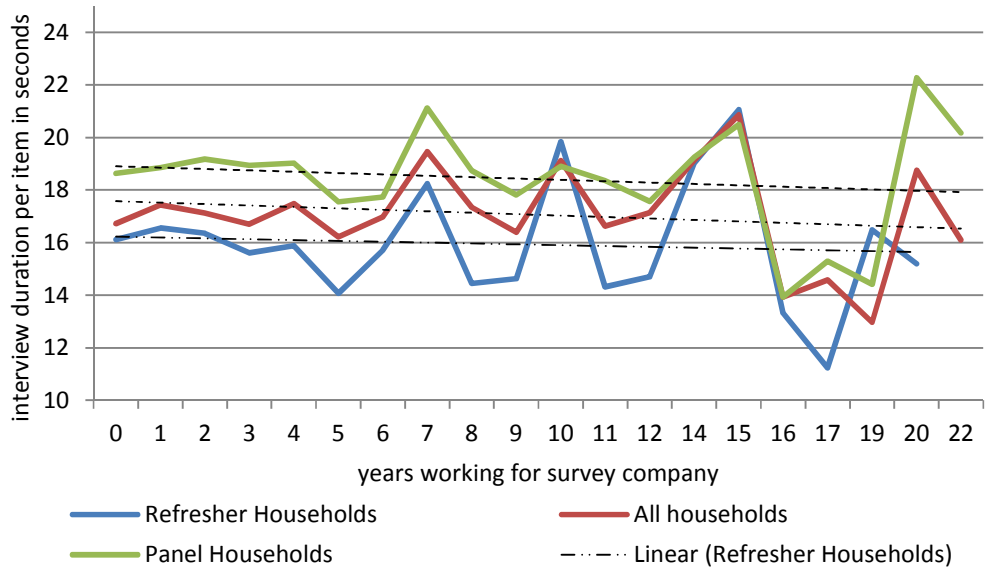
Total interview times (not controlling for the different number of questions) are shorter for panel than for refresher interviewers, but this could just be the result of the intended matching of panel interviewers with panel households, i.e. the share of panel households interviewed by a given panel interviewer is higher and the overall interview duration of panel households is shorter, as we will show below. This difference in structure could also be the reason for the observed equality of mean interview duration per item. We will show below that interview duration per item is longer for panel households than for refresher households.

We continue the presentation of results with the two indicators of experience most commonly used in the literature, ie. the experience in terms of years working as an interviewer and the within wave experiences measured by the number of interviews in a sequence of interviews.

We find a small downward trend for interview duration per item as the experience of the interviewer in terms of years worked increases, for all households as well as for the two different groups of households. When interpreting these results one has to keep in mind, that in order for an interviewer to be a panel interviewer, the interviewer has to have worked for the survey company for at least four years. This means that the duration measurements for panel and refresher households for interviewers with little experience (less than 4 years), is exclusively based on non-panel interviewers. We will show below that panel interviewers are on average quicker than non-panel interviewers in interviewing both types of households. This structure may contribute to the downward trend observed in Figure 3, indicating an overall reduction in mean interview duration per item with increasing experience for all interviewers, regardless of the type of households being interviewed.

To check whether there is a significant decline of interview duration per item, we regress this measure on interview experience and a constant. If all households are considered the negative trend is insignificant. If the sample is split into refresher and panel households the trend is negative and significant in both subsamples. It shows that analysing all types of experience together can be important for understanding the drivers of interview duration. In any case, the coefficient estimate on years of experience is about -0.06,

indicating that one extra year of experience reduces the interview duration only marginally.

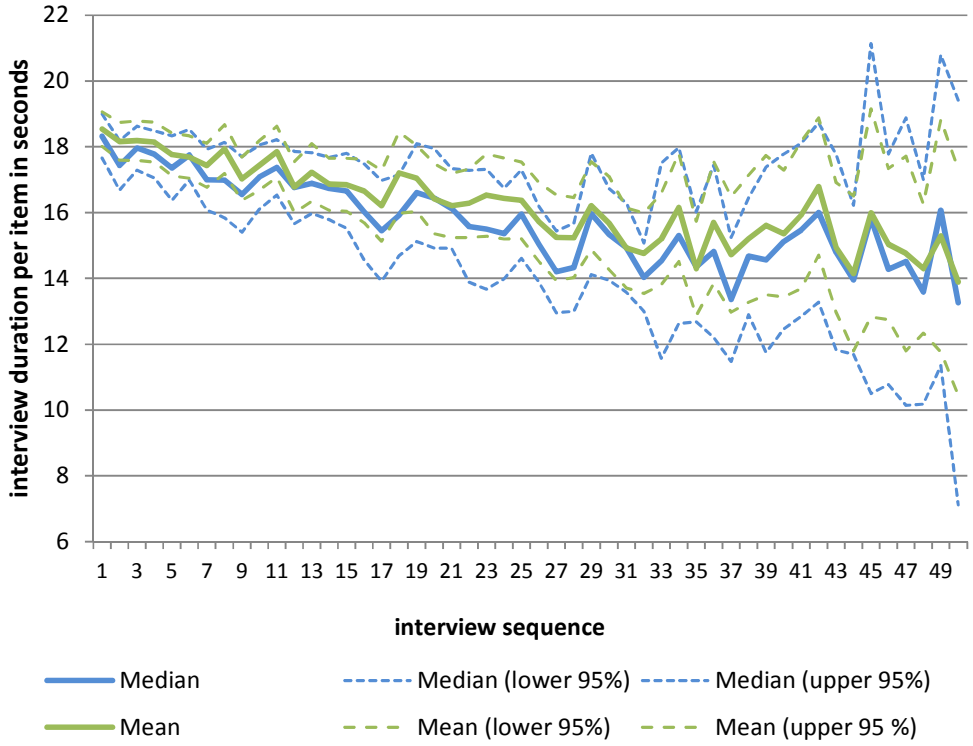


Notes: Zero years of experience indicates that the interviewer just started working for the survey company in the year the third wave of the PHF survey started.

Figure 3: Average Interview Duration per Item, by Years of Experience

For the experience measure based on the interview sequence we conduct a similar analysis. In Figure 4 below we show that for this measure a clear downward trend exists for most interview sequences. Interestingly this downward trend can be observed for interviewers with short and long interview sequences (see figure 6 in the appendix for details).[9] Even if the interviewer only conducts five interviews we already see a downward trending median interview duration per item. A regression model with a constant and the interview sequence as an explanatory variable yields a significant coefficient estimate of -0.027 for interviewer sequence in the model for median interview duration per item and -0.033 for the regression with mean interview duration per item as a dependent variable. Again the impact seems to be small. However, the coefficient is only about half the size of that for years of experience with the survey company. Put differently, two additional interviews have on average the same effect as one more year of experience working as an interviewer. The numbers presented are averages over the whole range of the interview sequences. If we restrict the analysis to interview sequences of up to 10, 20, 30 or 40 interviews, we see that the decline in average and median interview time is strongest in the beginning and then levels off. The difference in mean between interview one and ten is more than one second, between 10 and 20 again more than one second and then only 0.7 seconds between 20 and 30, and less than half a second between interviews 30 and 40.

[9] A figure with the median interview durations per item by interview sequence for interviewers with different numbers of interviews is included in the appendix.



Notes: Dashed lines are 95% confidence intervals for the mean and median, respectively. The graph is cut-off at interview sequences up to 50 interviews.

Figure 4: Median and Average Interview Duration per Item, by Interview Sequence.

4.1.2 Households' Experience

Panel households have on average shorter interviews than refresher households. However, they also answer fewer questions, due to the filtering of the questionnaire. If this effect is taken into account we see that per item the interview is on average significantly longer for panel households than for refresher households. The same is true for the median. The difference is almost 3 seconds, ie. about 20% of the mean interview duration per item.

Table 3: Average Interview Duration per Item, by Households' Experience with PHF

	Number of households	Mean interview duration	Median interview duration per item	Mean interview duration per item	Standard deviation
Panel households	1,911	62.9	18.1	18.4	4.7
Refresher households	2,182	67.3	15.3	15.8	4.1
All households	4,093	65.3	16.5	17	4.6

4.1.3 Combinations of Interviewers' and Respondents' Experience

So far we have looked at the different types of experience independently. In this subsection we now present results for combinations of interviewer and respondent experience. We restrict the analysis to one particular combination, ie. interviewers' and respondents' panel experience. An analysis including more types of experience at the same time, while also controlling for interviewer and respondent characteristics, will be presented in the next section. We have already seen that panel respondents take longer per item to answer the questionnaire and that on average the interview duration per item of panel and refresher interviewers are the same. Table 4 below clearly shows that the latter finding is due to the different share of panel respondents interviewed by panel interviewers compared with new interviewers. Panel interviewers are indeed quicker on average for interviews with both types of households in wave 2. The difference in interview duration per item is about one second and statistically significant. This analysis indicates that it is important to take both the respondents' and the interviewers' experience into account, as well as the composition of the sample.
.

Table 4: Average Interview Duration per Item, by Interviewers' and Households' Experience with PHF study

Mean interview duration per item			
	HH: Panel	HH: Refresher	ttest – panel vs refresher within interviewer group
Panel Interviewer	18.1 (1,301)	15.1 (689)	***
Interviewer new to PHF	19.0 (610)	16.1 (1,493)	***
ttest – panel vs new int. within household group	***	***	(4,093)
Mean interview duration			
	HH: Panel	HH: Refresher	ttest – panel vs refresher within interviewer group
Panel Interviewer	61.3 (1,301)	63.7 (689)	***
Interviewer new to PHF	66.4 (610)	69.0 (1,493)	***
ttest – panel vs new int. within hh group	***	***	(4,093)

Notes: number of observations in each cell shown in brackets. *** 99% significance level.

4.2 Multi-Level Analysis

The descriptive analysis above has shown that taking a multi-dimension view is necessary to understand the mechanisms behind interview duration per item. We have shown that the disproportionate assignment of panel households to panel interviewers can matter for the interpretation of the findings. In addition, different types of experience seem to matter for interview duration. In this section we thus implement a more comprehensive approach and estimate multi-level regression models, which allow us to analyse the relationship between interviewer and respondent experience and characteristics with interview duration, controlling for interviewer and respondent characteristics.
We set up a two-level model, where respondents are nested within interviewers. The model contains the following variables at the interview level:

- years of working at survey company
- indicator for experience as PHF interviewer ("panel interviewer")
- age and gender of the interviewer
- total number of interviews in wave 2
- position of interview in interviewers sequence within wave 2 (as such and squared term)

and at the respondent level:

- indicator for panel households
- age, gender and education of the respondent
- household size and estimate of household wealth

Additionally we control for the para-data indicators described in section 3.3. Since para-data is not available for 111 households, the multi-level analysis is thus based on only 3,982 observations, instead of the 4,093 used in the descriptive part of the paper.[10]
The results of the multi-level estimations are presented in Table 5. One result across all models is that respondents' experience does matter and matters more than interviewers' experience.[11] With respect to the former we find a large and significant value for the status of being a panel respondent. In terms of interviewers' experience only within study experience seems to matter. [12] For the interviewers' experience measured as the number of years they have been working for the survey company, we estimate a negative coefficient similar to the one from the descriptive analysis, but fail to reach conventional levels of significance. For the panel status of the interviewer we even find a positive yet insignificant effect. The interviewers work experience and the panel status of the interviewer is by definition highly correlated. A panel interviewer has to have worked for at least 3 years at the survey company, since the gap between wave one and wave two was three years. Indeed, the correlation coefficient between the two variables is almost 70%. We therefore test for the joint significance of the two indicators and run two

[10] Dropping the 111 households does not alter the descriptive results.
[11] In general, respondent characteristics are more important than interviewer characteristics. This can be seen from the R2 estimates in Table 5.
[12] The significance of the squared term for the interview sequence indicates that the interview time per item decreases in a non-linear fashion, until a value of 75 interviews the impact is negative and after that it turns positive. However, we have only one percent of interviewers with more than 75 interviews and an even smaller share of households that are number 76 or more in the interviewer sequence.

robustness checks including only the panel status or the number of years working as an interviewer in the model, respectively. We have to reject joint significance and the robustness checks show that the individual indicators are insignificant. This finding is consistent with the (unconditional) descriptive result on the interviewer experience presented in table 2 above. The estimation results are, however, not consistent with the findings in table 4, that panel interviewers are quicker in interviewing refresher and panel households. To investigate this issue further, we look at combinations of interviewer and respondent experience directly (model 6). We find, consistent with table 4 that panel respondents have longer per item interviews than refresher respondents, regardless of whether the interview is done by a new or a panel interviewer. However, the difference between the groups "new household and new interviewer" versus "new household and panel interviewer" is insignificant and the coefficients estimated for the groups "panel household and new interviewer" versus "panel household and panel interviewer" are not different. Given that not even in the model without any variables but the interviewer experience the panel status of the interviewer is insignificant, we argue that this type of experience is not related to interview duration per item.[13]

The ICC estimates show that interviewer behaviour is important for understanding interview duration. More than 25% of the variance in interview duration can be attributed to interviewers.

Concerning the socio-demographic variables related to the respondents we find, that the respondents' characteristics are more important than the interviewers. Respondent's age and gender are significantly related to interview duration. As documented in the literature, the interview duration per item increases with age. Women seem to have longer per item interviews than men, but surprisingly, this relationship becomes insignificant one we control for the additional para-data collected on the interest of respondents in the survey, use of documents, etc.. It looks like female respondent differ with respect to the para-data so that if the para-data is not included the female variable picks up those differences. Looking at the para-data for male and female FKP separately we find only small differences with respect to the use of documents. However, the interest in the survey and the ability to express themselves is rated lower by the interviewers for female than male respondents. The interviewers also report that more female respondents (24%) than male respondents (18%) are suspicious about the interview, and more females had at least minor difficulties in answer the questionnaire than man.

[13] Further analysis shows that in a simple linear regression framework we can obtain results similar to the one in table 4, if we only include the indicator combining the panel status of the household with the panel status of the interviewer. The estimated coefficients are in line with the descriptive results of a 1 second difference between panel and new interviewers. This result holds even if we cluster standard errors at the interview level. As soon as we add other control variables the coefficient is no longer significant. Regression tables for those robustness checks are available upon request.

Table 5: Multi-Level Estimation - Dependent Variable: Interview Duration per Item

		Model 1: Constant only		Model 2: + interviewer exp.		Model 3: + respondent experience		Model 4: + int and hh characteristics		Model 5: + para-data		Model 6: interaction	
		Coef.	se	Coef.	se	coefficient	se	coefficient	se	coefficient	se	coefficient	se
Experience	Interviewer's experience at survey company			-0.076	0.0545	-0.069	0.052	-0.089	0.053	-0.079	0.053	-0.081	-0.053
	Panel Interviewer			0.834	0.468	0.114	0.452	0.264	0.470	0.189	0.475		
	Interview seq					-0.0863***	0.010	-0.0667***	0.010	-0.0542***	0.010	-0.0547***	-0.010
	Interview seq^2					0.001***	0.000	0.000***	0.000	0.000**	0.000	0.000**	0.000
	Panel household					2.642***	0.157	2.575***	0.155	2.625***	0.155		
	Panel interviewer - new household											0.403	-0.496
	New Interviewer - panel household											2.864***	-0.221
	Panel Interviewer - panel household											2.827***	-0.479
Respondent	age							0.0403***	0.004	0.0380***	0.004	0.0381***	-0.004
	female							0.334**	0.123	0.220	0.121	0.223	-0.121
	education level 2							-0.723	0.741	-0.696	0.721	-0.715	-0.721
	education level 3							-0.712	0.708	-0.830	0.693	-0.845	-0.693
	education level 4							-0.833	0.733	-1.087	0.720	-1.099	-0.720
	education level 5							-0.656	0.710	-0.879	0.698	-0.900	-0.698
	education level 6							-0.889	0.771	-1.215	0.759	-1.230	-0.759
Household	hh size 2							0.139	0.154	0.044	0.150	0.048	-0.150
	hh size 3							-0.494*	0.202	-0.627**	0.197	-0.621**	-0.197
	hh size 4							-0.593**	0.221	-0.774***	0.216	-0.766***	-0.216
	hh size 5+							-0.906**	0.307	-1.043***	0.299	-1.037***	-0.299
	wealth (estimated)							0.000	0.000	0.000	0.000	0.000	0.000
Interviewer	female							0.543	0.336	0.523	0.340	0.533	-0.340

	age							-0.005	0.016	-0.006	0.016	-0.006	-0.016
	total number of interviews in W2							-0.012	0.010	-0.014	0.010	-0.014	-0.010
Paradata	Respondent suspicious before interview									0.078	0.167	0.070	-0.167
	Difficulty in answering: minor									-0.634*	0.297	-0.647*	-0.297
	Difficulty in answering: small									-0.942***	0.284	-0.946***	-0.284
	Difficulty in answering: none									-1.443***	0.311	-1.446***	-0.311
	Ability to express oneself: good									0.739**	0.281	0.725**	-0.281
	Ability to express oneself: very good									0.521	0.304	0.507	-0.304
	Interest in survey: above average									0.545***	0.157	0.539***	-0.157
	Interest in survey: very high									1.153***	0.209	1.145***	-0.209
	Use of documents: rarely									1.006***	0.167	1.008***	-0.167
	Use of doc.: sometimes									1.654***	0.169	1.659***	-0.169
	Use of documents: often									2.123***	0.261	2.138***	-0.261
	Constant	17.53***	0.161	17.57***	0.238	17.39***	0.244	16.12***	1.203	15.87***	1.220	15.80***	-1.221
	Var (interviewer)	5.205***	0.587	5.157***	0.581	4.653***	0.523	4.696***	0.523	4.908***	0.539	4.910***	-0.539
	Var (residuals)	15.70***	0.364	15.70***	0.364	13.80***	0.32	13.13***	0.304	12.38***	0.287	12.37***	-0.287
	ICC (Interviewers)	0.249		0.247		0.252		0.263		0.284		0.284	
	R2 interviewer			0.009		0.106		0.098		0.057		0.057	
	R2 residual			0.000		0.121		0.164		0.212		0.212	
	Observations	3982		3982		3982		3982		3982		3982	

Notes: Raudenbush and Bryk R2, *** 99% significance level, ** 95% significance level, * 90% significance level. Reference groups not shown.

4. Conclusions and Future Research

Our analysis shows that experience matters for the duration of an interview. We distinguish experience of respondents from experience of interviewers and look specifically at experience across waves of a panel study.

We find that respondents' within project experience, ie. being a panel respondent or not, correlates significantly with interview duration per item. As far as interviewers are concerned, we find that within study experience of interviewers, ie. the number of interviews conducted in the same wave prior to the given interview, is significantly correlated with interview duration. However, being a panel interviewer alone is not significantly related to interview duration per item.

The missing correlation between being a panel interviewer and interview duration per item may be a result of the long gap of three years between interviews in the German wealth survey. The interviewers visit the household only every three years and most of the panel interviewers have conducted many other interviews for other studies in between two waves of the PHF survey. The experience gained with the first wealth survey can thus be assumed to have decayed when the second wave starts. Furthermore, both refresher and panel interviewers get a similar pre-fieldwork training before each wave. Part of the training comprises conducting two test interviews and going through the CAPI program alone. The in-person training also focuses on the parts of the CAPI program which are more complex to administer. Both panel and new interviews can thus be expected to have a similar understanding with regards to the current wave of the survey. For panel households, the study is likely to be different from other studies they participate in, due to its length and topic. This may lead to a larger "experience" effect, simply because households remember the study even after a long time.

Surprisingly, we find that panel households have longer per item interviews than refresher households. To analyse the reasons and implications of this finding are beyond the scope of our study. An analysis of the response behaviour, e.g. item non-response rates, rounding or other quality indicators, may shed some light on these issues. These indicators may help to understand whether the panel households take longer because they answer questions with greater care or because they rush through the questionnaire and take short cuts. Other channels, like whether the panel respondents "chat" more with the interviewer or whether they simply are planning for longer interviews and thus don't have to finish quickly, are harder to explore. Recording interviews is not considered in the German wealth survey.

The number of the interview in the interviewer's sequence of interviews seems to matter. Also here the question arises why this is the case. Do interviewers learn how to administer the survey and are better prepared to answer questions in subsequent interviews or do they simply rush households through the survey and induce them to take short cuts as they learn, which short-cuts are most "beneficial"? Again, an analysis of the specific within interview response patterns may contribute to answering these questions.

In summary, our paper documents significant differences in interview duration per item for different types of interviewers and respondents, but remain silent about the specific processes behind the heterogeneous effects. We also leave an assessment of what the different interview times mean, e.g. in terms of data quality for future research. Another natural extension of our study is to take the "difficulty of each item into account. Despite the fact that we are confident that on average panel and refresher households do not differ markedly with respect to what types of questions they have to answer, it may be worth taking the specific question types posed to respondents into account (cf. Cooper and Kreuter, 2013) in future research.

Acknowledgements

We would like to thank Panagiota Tzamourani for helpful comments.

References

Bassili; J. N. 1996. The how and the why of response latency measurement in telephone surveys, in: Schwarz, N. and Sudman, S., Answering Questions: methodology for Determining Cognitive and Communicative Processesin Survey Research, 319-346, Jossey-Bass, San Francisco.

Cannell, C. F., K. H. Marquise and A. Laurent 1977. A Summary of Studies of Interviewing Methodology, Vital Health Stat 2(69), 1-78.

Couper, M. P. and F. Kreuter 2013. Using Paradata to Explore Item Level Response Times in Surveys, Journal of the Royal Statistical Society: Series A 176(1), 271-286.

Draisma, S. and Dijkstra, W. 2004. Response Latency and (para) linguistic expressions as indicators of response error, in: Presser, S., Rogthgeb, J., cooper, M., Lesser, J., Martin, E. Martin, J. and Singer, E., Methods for Testing and Evaluating Survey Questionnaires, 131-147, Wiley, Hoboken.

Fowler, F. J. 1991. Reducing Interviewer-Related Error Through Interviewer Training, Supervision and Other Means, in: P. Biemer, R.M. Groves, L. Lyberg, N. A. Mathiowetz, and S. Sudan, Measurement Errors in Surveys. Wiley, Ney york

Gummer, T. and J. Roßmann 2015. Explaining Interview Duration iin Web Surveys: A Multilevel Approach, Social Science Computer Review 33(2), 217-234.

Heerwegh, D. and G. Loosveldt 2002. An evaluation of the effect of response format on data quality in Web surveys, Social Science Computer Review 20, 471-484.

Loosveldt, G. and K. Beullens 2013. How Long will it Take? An Analysis of Interviewer Length in the Fifth Round of the European Social Survey, Survey Research Methods 7(2), 69-78.

Mayerl, J. 2013. Response Latency Measurement in Surveys. Detecting Strong Attitudes and Response Effects, Survey Methods: Insights from the Field, retrieved from hhtp://surveyinsights.org/?p=1063, doi:10.13094/SMIF-2013-00005

Olson K. and I. Bilgen 2011. The Role of Interviewer Experience on Acquiescence, Public Opinion Quarterly 75(1), 99-114.

Olson, K. and A. Peytchev 2007. Effects of Interviewer Experience on Interview Pace and Interviewer Attitudes, Public Opinion Quarterly 71(2), 273-286.

Pickery, J. and G. Loosveldt 2001. An Exploration of Question Characteristics the Mediate Interviewer Effects on Item Nonresponse, Journal of Official Statistics 17, 337-350.

Stocké, V. and B. Langfeldt 2004. Effects of Survey Experience on Respondents' Attitudes Towards Surveys, Bulletin of Sociological Methodology 81, available online at (accessed: Oct 1, 2016): http://bms.revues.org/1094.

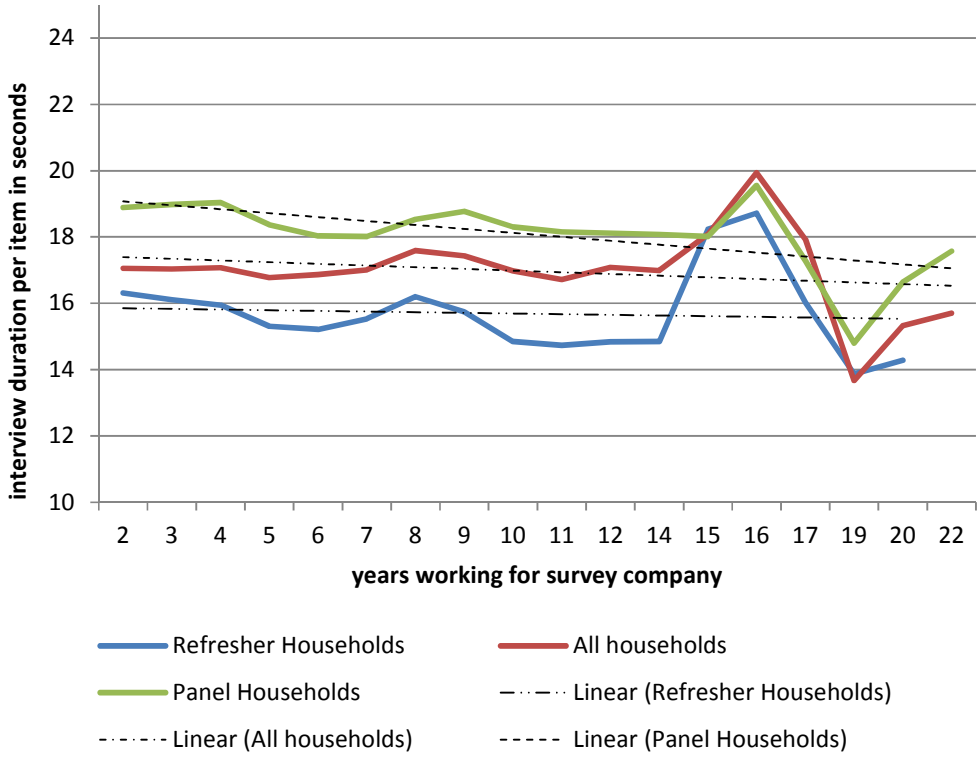
Toepoel, V., M. Das, A. van Soest 2008. Effects of Design in Web Surveys. Comparing Trained and Fresh Respondents, Public Opinion Quarterly 72(5), 985-1007, doi: https://doi.org/10.1093/poq/nfn060.

Wuyst, C., C. Vandenplas and G. Loosveldt 2018. An Interviewer-Oriented Analysis of Interview Speed: Experience, Burden or Both?, unpublished manuscript.

Yan, T. and R. Tourangeau 2008. Fast and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Surveys Response Times, Applied Cognitive Psychology 22, 51-68, doi: 10.1002/acp.1331.

Yan, T., L. Ryan, S. E. Becker and J. Smith 2015. Assessing the Answers to a Global Subjective Well-being Question Through Response Time, Survey Research Methods 9(2), 101-109

Appendix



Notes: Moving averages for interview duration per item, combing the current experience bin with the previous two.

Figure 5: Average Interview Duration per Item (Moving Averages), by Interviewers' Years of Experience.

Figure 6: Median Interview Duration per Item, by Interview Sequence

Table 6: Interview Duration, by Selected Household and Interviewer Characteristics

	Interview duration (min)	Iterview duration per item (sec)
FKP: female	64.6	17.1
FKP: male	65.7 *	16.9
FKP: age < 50	65.6	16.1
FKP: age 50+	65.0	17.5 ***
Int: Female	66.6	17.4
Int: Male	64.5 ***	16.8 ***
Int: age <50	67.4	17.5
Int: age 50+	64.9 ***	16.9 ***

.

Table 7: Summary Statistics for Household Characteristics, Interviewer Characteristics and Para-Data

Variable	scale	mean	sd
Interviewer Characteristics			
female	dummy	0.35	0.48
age	continuous	59.87	9.36
total number of interviews in W2	count data	31.82	26.22
experience at survey company	continuous	5.10	4.15
panel interviewer	dummy	0.49	0.50
Interviewer sequence	count data	16.63	17.74
FKP/Household Characteristics			
panel household	dummy	0.47	0.50
age	continuous	55.01	16.02
female	dummy	0.43	0.50
education level 2	dummy	0.06	0.24
education level 3	dummy	0.42	0.49
education level 4	dummy	0.08	0.27
education level 5	dummy	0.40	0.49
education level 6	dummy	0.04	0.19
household size	continuous	2.30	1.14
net wealth (estimated)	continuous, in thousand euros	289.53	949.38
Para-data			
Respondent suspicious before interview	dummy	0.21	0.40
Difficulty in answering: minor	dummy	0.12	0.33
Difficulty in answering: little	dummy	0.45	0.50
Difficulty in answering: none	dummy	0.37	0.48
Ability to express oneself: good	dummy	0.38	0.49
Ability to express oneself: very good	dummy	0.55	0.50
Interest in survey: above average	dummy	0.41	0.49
Interest in survey: very high	dummy	0.23	0.42
Use of documents: rarely	dummy	0.16	0.36
Use of documents: sometimes	dummy	0.17	0.37
Use of documents: often	dummy	0.06	0.23