PAPI is gone, but errors remain: Non-sampling errors in household surveys

by Mark Brooks, Rattiya S. Lippe, and Hermann Waibel

PAPI is gone, but errors remain

– Non-sampling errors in household surveys –

Mark Brooks[a, b], Rattiya S. Lippe[c], Hermann Waibel[a]

June 2021

Abstract

Reliable and accurate data are key to the development and implementation of suitable policies. In recent years, the overall quality of data drawn from household surveys, which are common sources of data in developing countries, has been found to be lacking. In this study, we examine the effects of respondent and interviewer characteristics, their interactions and the interview and survey environment on the count of non-sampling error. Further, we examine whether findings are consistent across two countries with differing cultural backgrounds and interviewer profiles. Using data from the 2017 wave of a household panel survey in Thailand and Vietnam, we model the influence of these determinants by applying a negative binomial regression both with a combined sample and at the country-level. One regression variant is applied for each of the three identified types of non-sampling error: (a) item non-response due to missing data, (b) item non-response due to refusal, and (c) measurement error. We show that determinants of non-sampling error stem not only from respondent and interviewer

a Faculty of Economics and Management, Institute of Development and Agricultural Economics, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany

b Corresponding author, Tel.: +49 511-762-5414, Fax: +49 511 -762-2667, E-mail: brooks@ifgb.uni- hannover.de

c Thuenen Institute of International Forestry and Forest Economics, Leuschnerstraße 91, 21031 Hamburg-Bergedorf, Germany

characteristics, but also from characteristics of the interview and survey environment such as the timeframe in which an interview takes place. Notably, our study finds find that differing cultural contexts, congruency of respondent-interviewer characteristics such as ethnicity and personality traits of interviewers must be accounted for in order to improve the quality of household survey data in developing countries. Further, by comparing our results with a previous study analysing determinants of non-sampling error using data from PAPI waves of the same survey, we are able to show that when CAPI is implemented successfully, item non-response is essentially eliminated. Measurement error, however, remains a persistent threat that must be addressed by survey providers.

Keywords: Non-sampling error, Data quality, Paradata, Household survey, Southeast Asia, TVSEP

JEL: C81, C83, O10

1. Introduction

In the context of developing countries, demographic and socio-economic data, which are key for developing and monitoring development policies, are often sourced from household surveys. The presence of non-sampling errors such as missing data and measurement error constitutes a substantial threat to household survey data quality and confronts researchers and policymakers with a degree of uncertainty. Shortcomings of such data negatively influence their representativeness and applicability for policy formulation. Especially in developing countries, the quantity of high-quality household panel survey data remains sparse (Dang & Carletto, 2018). For example, Booth (2019) and Gibson (2016) identify that although in recent years an abundance of household survey data is available for many countries, relatively few are suitable for calculating reliable poverty estimates. Accordingly, improving the reliability and validity of household survey data is a necessary complementary task in order to ensure that important policy recommendations can be deduced.

Innovations to survey tools utilized in household surveys in developing countries such as Computer Assisted Personal Interviews (CAPI) have increased the overall effectiveness of surveys and the quality of their outputs. The World Bank, for example, by designing and implementing the android-based software "Survey Solutions" has made an important contribution to ensure that survey providers in developing countries are able to develop the capacities needed to procure high-quality data in a timely manner. Experimental evidence suggests that conducting surveys with CAPI, over the more traditional Paper and Pencil Interviews (PAPI) commonly used to conduct surveys in developing countries, can prevent numerous errors. The implementation of, for example, automated routing and plausibility checks incorporated in computerized questionnaires prevents missing data and assists in pre-emptively identifying potential measurement errors (Caeyers et al., 2012; de Leeuw et al., 1995). In spite of advances made in terms of data access, use, and collection, innovations to CAPI do not automatically solve the persistent problem of low-quality data (Meyer et al., 2015). Errors arising from observable interviewer characteristics, misinterpretation of questions, and difficult interview and survey conditions remain a challenge, especially in developing countries (Lupu & Michelitch, 2018).

We identified at least four research gaps in the literature on non-sampling errors concerning household surveys in developing countries. Firstly, most studies rely on either cross-sectional or experimental data and are therefore limited in scope as regards the type of non-sampling errors studied. Secondly, emphasis has been placed on quantitative interviewer and respondent characteristics such as age, gender, and education, whereas qualitative information such as interviewer and respondent behaviours, their personality traits, and motivations are not.

Thirdly, most studies focus on individual determinants of non-sampling errors such as the effects of the interviewer and/or respondent characteristics, but rarely account for the circumstances of the interview or survey itself (Lupu & Michelitch, 2018). Finally, survey paradata[1] are infrequently available to researchers and survey providers rarely supply extensive albeit useful, supplementary paradata. In summary, what is missing in the literature so far, is a comprehensive approach that takes quantitative and qualitative information simultaneously into account and thereby allows the identification of the relative importance of: (a) interviewer characteristics, (b) respondent characteristics, (c) the interview and survey environment and (d) time.

The remainder of the paper is structured as follows: Section 2 briefly defines non-sampling errors and provides an overview of determinants of non-sampling errors identified in the literature. Thus, laying the foundation for formulating the empirical model to estimate factors influencing the prevalence of non-sampling errors presented in Section 3. Section 4 describes the study area, the survey implementation and survey data. Section 5 summarizes and discusses the main results. In the final section, conclusions are drawn and recommendations toward improving household survey data quality are offered.

2. Literature review

According to the literature on data quality (e.g. Groves, 1989; Weisberg, 2005), non-sampling errors consist of (i) coverage error, (ii) non-response errors, and (iii) measurement error. As this paper focuses on non-sampling errors that occur throughout the interview process, we omit coverage errors.

1 Paradata refer to data collected that describe the process of survey production that are not part of the interview itself (Kreuter et al., 2010)

Non-response occurs when measurements for a sampling unit or for a specific item of the survey instrument cannot be obtained. Thus, non-response affects the representativeness of data and its ability to be externalized for policy formulation. The literature differentiates between unit- and item non-response. Unit non-response refers to sampling units that are entirely missing from the survey data base (e.g. the interviewer could not contact or locate a household) and is hence omitted from the analysis. Item non-response, in contrast, occurs when responses of the sampling unit are only available partially. Questions may be erroneously skipped and respondents unable to or unwilling to provide essential information, in particular on sensitive subjects (Lynn & Clarke, 2002). A deviation of the value provided by the sample unit from its true value constitutes a measurement error. There are three types of measurement error: response, interviewer and postsurvey error (Weisberg, 2005). Deviation from the true value of response on behalf of the respondent, for example, due to only vaguely recalling fertilizer quantities used on their rice plot and thus providing an estimated figure, constitutes a response error. An interviewer error occurs when a question or response is inaccurately modified in such a way that its meaning changes. For example, rephrasing "how healthy do you feel?" as "are you doing okay?" changes the original intent of the survey item and may provoke a different response. Postsurvey error frequently result from data entry, merging, and processing. For example, in PAPI surveys questionnaires are entered manually to a database and processed by survey staff, thereby potentially introducing typographical errors (Glewwe & Dang, 2008). While such errors can be bypassed in other survey modes such as CAPI, which no longer require transcription of field interviews to a database, other forms of post-survey error remain due to faulty implementation of merge commands, erroneous syntax or invalid plausibility rules that are applied during data cleaning.

Sources of non-sampling error range from individual characteristics of respondents and interviewers, their interactions, and characteristics of the interview and survey environment.

An overview of key findings of the literature on determinants of non-sampling error is provided in the following paragraphs:

Firstly, non-sampling errors arise during the process of respondents formulating their response to individual survey items. Responses can deviate from their true value due to, for example, misinterpretations of question intent, misunderstanding of the question itself, or deliberate misreporting (Tourangeau et al., 2000). Additionally, elements of survey design such as deviations in the period of recall influences the prevalence of non-sampling error. Long periods of recall are found to introduce less accurate representations of household activities with recall bias being particularly relevant for surveys that collect data on agricultural activities, which are seasonal in nature. Hence, recall bias is thought to increase the wider and further back the defined reference period of a survey is (e.g. Beegle et al., 2012; Wollburg et al., 2020). Another source of error lies in the selection of respondents in household surveys. The utilization of proxy respondents, who are probed in detail about characteristics and activities of other members of their household, is common in household surveys due to monetary constraints. Interviewing each individual household member is seldom feasible and hence data quality is traded off for more extensive household information (Bardasi et al., 2011). Survey providers in developing countries frequently advocate targeting household heads in proxy interviews as they are most knowledgeable about the household due to their high status within its hierarchy. Results from the literature are mixed with Phung et al. (2015), for example, finding that interviews with household heads are more complete (e.g. less item non-response) than interviews with other proxy respondents. Conversely, a study by Fisher et al. (2010) found that household heads significantly underestimate household and household member income, in particular that of their spouses. The cognitive ability of respondents has also been identified as a potential determinant of non-sampling error. Cognitive ability is proxied for by using the age

and level of education of respondents in studies, which find that elderly respondents and those with lower education (e.g. lower cognitive ability) produce data of significantly lower quality (Knäuper et al.,1997; Krosnick, 1991). Further, their findings suggest that such respondents are likely to satisfice more frequently during interviews. Satisficing refers to respondents actively deciding to forgo steps and/or utilizing shortcuts in order to minimize effort required in formulating their response, thus their goal is to provide satisfactory rather than optimal responses (Krosnick & Alwin, 1987). For example, respondents may opt to straight-line (e.g. always choose the first response option), answer at random, or select options such as "don't know" (Barge & Gehlbach, 2012).

Secondly, deviations from designated survey procedures on behalf of interviewers can influence the respondent in their response formulation. Prominent examples of behaviours of the interviewer that influence responses are rephrasing questions, neglecting to follow interview instructions, skipping questions to reduce workload (e.g. satisficing) or due to perceived sensitivity of a subject. On the one hand, measurement errors may occur during data entry if interviewers incorrectly record a response or make a typographical error. On the other hand, directly assisting the respondent in framing their response, either by rephrasing difficult questions or explaining question meaning alongside inappropriate probing techniques to elicit responses, can lead to measurement error. For example, variations in interviewer speech in the form of emphasis or intonation of question text is argued to influence respondents and potentially introduce measurement error (Biemer, 2010). Such faulty methods of enumeration are frequently a result of insufficient training or lack of experience in conducting interviews (Campanelli & O`Muircheartaigh, 1999; Singer et al., 1983). Prior experience of interviewers in survey activities can provide a well-grounded foundation of knowledge on interviewing behaviours that elicit cooperation, which is a prerequisite for obtaining accurate responses

(Couper & Groves, 1992). Accordingly, well-trained and experienced interviewers are generally thought to procure higher quality interviews (Olson & Bilgen, 2011). However, interviewers without an extensive background in survey labour have in some cases been found to provide data that is of higher quality. For example, Fowler and Mangione (1990) observe that inexperienced interviewers are less likely to be biased by prior experience in other survey contexts and are thus more likely to follow closely existing survey guidelines, which in turn improves the quality of their collected data. The literature has attempted to determine whether or not there is an effect of interviewer gender on the quality of collected data, but to date the results are inconclusive. While authors such as Campanelli and O'Muircheartaigh (1999) find that male interviewers collect data of higher quality, Phung et al. (2015) observe the opposite with female interviewers providing interviews with a lower count of non-sampling error than their male counterparts. In terms of quantitative interviewer characteristics, interviewers exerting friendly or motivating behaviours were found to procure higher rates of cooperation (Jäckle et al., 2013; Olson et al., 2016).

Thirdly, congruency of interviewer and respondent characteristics has been examined in the literature based on the premise that an interview is a structured social interaction in which demographic and socio-economic characteristics of participants interact with one another (Kahn & Cannell, 1957). Congruency of characteristics such as age, gender and ethnicity are of interest (Baird et al., 2008; Feskens et al., 2006; Phung et al., 2015), in particular in surveys that deal with sensitive topics (Catania et al., 1996).

Finally, the survey and interview environment are found to be drivers of non-sampling error. Studies on survey data quality often control for the duration of interviews and find that longer interviews frequently result in interviews that are prone to non-sampling error. This loss of quality is argued to stem from increasing levels of interviewer and respondent fatigue alongside

potential loss of motivation in drawn-out interviews (Galesic & Bosnjak, 2009; Phung et al., 2015). The presence of others, in particular during face-to-face interviews, is less frequently accounted for and provides an incentive for respondents to adjust their responses to adhere to perceived social norms (Krumpal, 2013). Further, the quantity and quality of supervision is found to significantly influence data quality. For example, inadequate and untimely scrutiny of data and data processing can lead to underlying issues in the survey instrument or undesirable behaviours of interviewers not being identified resulting in preventable errors (Groves et al., 2011).

3. Methodology

Based on the literature review, we hypothesize that non-sampling errors in their various forms are influenced by (i) interviewer characteristics, (ii) respondent characteristics, (iii) congruency of aforementioned characteristics, (iv) the interview and survey environment.

In terms of the types of non-sampling error considered, we focus on item non-response and measurement error. Additionally, we disaggregate item non-response in order to differentiate between two types of non-response. Firstly, we consider missing data, which represents cases in which a response in not entered in the questionnaire due to an entry error on behalf of the interviewer or imperfect implementation of the computerized questionnaire. Secondly, we consider refusals, which are hypothesized to be driven by the respondent and their interaction with the interviewer. Refusals are identified in the context of the survey instrument through the application of the code “no answer”, which signals that the respondent was unwilling to provide an answer. Measurement errors, which are forthwith coined as implausible values are considered when a response does not comply with existing survey plausibility rules, answers are inconsistent throughout the interview, or for extreme outlier values that cannot be plausibly

explained by supplementary data, such as local market prices or responses from previous waves, or through call-backs to the household in question.

A count model in the form of a negative binomial regression is fitted to analyse the determinants of non-sampling error. Such models have previously been used to predict software faults (Yu, 2012) and to analyse the effects of interviewer and respondent characteristics on the quality of open-ended questions (Barth & Schmitz, 2021). A negative binomial regression was selected over other count models as our dependent variables are overdispersed with variance exceeding the mean. Further, likelihood-ratio tests were run in Stata in which the negative binomial regression was tested against Poisson and zero-inflated count models and found to be highly significant (Figure A1 & Figures A2-A4). Therefore, the negative binomial regression model was shown to outperform other count models and deemed preferable for our analysis of the determinants of non-sampling error. As interviews are not homogeneous, there are varying degrees of risk of non-sampling errors occurring. In the negative binomial regression model, we control for this by adding the overall number of survey items asked, which varies by interview based on household characteristics, as an exposure variable. Additionally, we control for the complexity of interviews by adding variables on the composition of the household and household activities such as household size, size of agricultural land and yearly per capita income.

We first establish a negative binomial regression model that encompasses the combined sample of households from Thailand and Vietnam. The dependent variable is accordingly the count of the non-sampling errors within an individual interview. Hereby, three variants are considered – one for each type of non-sampling error considered and the model is specified as follows:

$$\mu_i = \exp(\beta_k X_{ki} + \delta_m Z_{mi} + \rho_n F_{ni} + \eta_o I_{oi} + \vartheta_p S_{pi} + \varepsilon_i) \quad (1)$$

where μ_i are (a) the count of missing data; (b) refusals; and (c) implausible values in survey items for the interview of household i (i = 1, …, N), respectively. X_{ki} are respondent characteristics; Z_{mi} are interviewer characteristics; F_{oi} are congruent characteristics between the interviewer and respondent; I_{oi} are characteristics of the interview and survey environment; and S_{pi} are household characteristics.

In the next step we modify model (1) in order to analyse the determinants of non-sampling errors individually for our two countries c (c = 0, 1). The goal is to determine whether determinants are consistent in terms of their effect on the prevalence of non-sampling errors. Further, we examine their applicability in two cultural contexts in order to determine whether survey providers must themselves identify key determinants of survey error on a country basis or whether results are applicable to a broad spectrum of survey backgrounds. The country-level specifications of the negative binomial regression model variants are as follows:

$$\mu_{ci} = \exp\left(\beta_k X_{kci} + \delta_m Z_{mci} + \rho_n F_{nci} + \eta_o I_{oi} + \vartheta_p S_{pci} + \varepsilon_{ci}\right) \quad (2)$$

While the categories from which independent variables are drawn in the country-level regression model are identical to those of the combined model variants, modifications are made based on differences between the two countries. Firstly, the populations of Thailand and Vietnam differ greatly in terms of ethnic diversity. Hence, cultural differences are more pronounced in Vietnam and interactions between different ethnic backgrounds are identified as a major cultural issue that must be accounted for (Dang, 2012). In addition, while the characteristics and backgrounds of interviewers are fairly homogeneous within each country, the backgrounds of Thai and Vietnamese interviewers differ significantly when compared to one another. Thus, we hypothesize that additional important insights on the role of the interviewer can be generated by disaggregating at the country-level. Finally, separating the data

base by country allows us to investigate potential provincial effects, which could be driven either by different conditions experienced in the province or team effects.

The variables included in our model variants as well as an overview of expected signs are presented in Table 1.

In terms of the missing model, we hypothesize that included characteristics of the respondent will not be significant. Accordingly, we control only for basic characteristics of the respondent such as age, gender, education and ethnicity in the missing data model variant. We argue that in CAPI the interviewer is solely responsible for data entry and hence interviewer characteristics will be significant drivers of missing data. We hypothesize that higher levels of experience in survey work and education will reduce the count of missing data. Additionally, stemming from a field of study that matches the survey topic or being local to the survey area is hypothesized to reduce the count of missing data. Further, we hypothesize that characteristics of the interview and survey environment such as the timing of the interview, speed of entry, technical malfunctions and increasing experience in the use of the computerized questionnaire as proxied by the variable survey weeks will significantly affect the prevalence of missing data.

Refusals and implausible values are hypothesized to be mainly driven by respondent and interviewer characteristics as well as their interactions (Baird et al., 2008; Phung et al., 2015). For example, we hypothesize that characteristics of respondents such as age (in proxy for cognitive ability), education and status within the household are significant drivers of non-sampling error. In terms of interviewer characteristics, we hypothesize that experience in survey work and higher education will reduce the count of refusal and implausible values. Further, personality traits such as openness, extraversion and agreeableness are hypothesized to significantly reduce the count thereof. The interview and survey environment are hypothesized as playing a lesser, albeit significant role. Longer interview durations and speedy

entries of responses are hypothesized to lead to increasing counts of non-sampling error. Additionally, the complexity of households as proxied by the household characteristics included in the models are hypothesized to significantly increase the count of missing data and implausible values.

Table 1. Description of explanatory variables and expected signs

Explanatory Variables	Description	Expected Sign (+/–/±/n.s.) (1) Missing	(2) Refusal	(3) Implausible	Source(s)
Respondent Characteristics					
Age	Age of respondent (years)	n.s.	±	+	Knäuper et al. (1997); Krosnick (1991)
Gender	1 if the respondent is male, 0 otherwise	n.s.	±	±	
Education	Respondent's highest education level (years)	n.s.	+	–	Knäuper et al. (1997); Krosnick (1991)
Ethnicity	1 if the respondent is from the country-level majority ethnic group of Thai|Kinh (TH|VN), 0 otherwise	n.s.	–	–	Feskens et al. (2016); Pennell et al. (2017)
Head of household	1 if the respondent is the head of the household, 0 otherwise		–	–	Phung et al. (2015)
Interviews	Number of times that the respondent has been interviewed		–	–	
Openness	Self-assessed openness (scale 1-7)			+	Jäckle et al. (2013); Olson et al. (2016)
Extraversion	Self-assessed extraversion (scale 1-7)		–		See above
Neuroticism	Self-assessed neuroticism (scale 1-7)		+		See above
Interviewer Characteristics					
Age	Age of interviewer (years)	+	±	+	
Gender	1 if the interviewer is male, 0 otherwise	–	±	–	Campanelli & O'Muircheartaigh (1999); Phung et al. (2015)
Education	Interviewer's highest education level (years)	–	–	–	
Ethnicity	1 if the interviewer is from the country-level majority ethnic group of Thai|Kinh (TH|VN), 0 otherwise	–	–	–	Feskens et al. (2016); Pennell et al. (2017)
Survey experience – Other	1 if the interviewer has prior experience in other surveys, 0 otherwise	–	–	–	Campanelli & O'Muircheartaigh (1999); Couper & Groves (1992); Singer et al. (1983)
Survey experience – TVSEP	1 if the interviewer has prior experience in TVSEP, 0 otherwise	–	–	–	See above
Years of survey experience	Interviewer's experience in survey work (years)	–	–	–	See above
Local	1 if interviewer was born in the province in which the survey is conducted, 0 otherwise	–	–	–	
Training performance	Aggregate weighted score of training performance indicators by sub-team leader (Scale 1-7)	–	–	–	

Explanatory Variables	Description	Expected Sign (+/–/±/n.s.)			Source(s)
		(1) Missing	(2) Refusal	(3) Implausible	
Openness	Weighted average of self-assessed openness and assessment of sub-team leader (scale 1-7)			+	Jäckle et al. (2013); Olson et al. (2016)
Extraversion	Weighted average of self-assessed extraversion and assessment of sub-team leader (scale 1-7)			–	See above
Agreeableness	Weighted average of self-assessed agreeableness and assessment of sub-team leader (scale 1-7)		–		See above
Field of study	1 if field of study is economics or agriculture 2 if field of study is sociology, languages or education; 3 if field of study is administration, politics or law.	–	–	–	
Interview/Survey Characteristics					
Interview duration	Duration of interview (minutes)	+	+	+	Galesic & Bosnjak (2009); Phung et al. (2015)
Answers per minute	Number of answers entered to tablet per minute	+	±	+	Couper & Hansen, (2002); Couper & Kreuter (2013); Kreuter et al. (2010)
First interview	1 if the interview took place during the morning, 0 otherwise	–	–	–	
Presence of others	1 if others aside from the interviewer and respondent were present during the interview, 0 otherwise	+	–	+	Krumpal (2013)
Tablet malfunction	1 if highly negative technical issues affected the interview (as assessed by the interviewer), 0 otherwise	+	+	+	
Survey week	Progression of the survey (weeks)	–	±	–	Baird et al. (2008); Townsend et al. (2013)
Country	1 if Thailand, 0 otherwise	±	±	±	
Household Characteristics					
HH size	Number of household members	+	+	+	
HH agricultural land size	Cumulative size of household agricultural land plots (1,000m²)	+	+	+	Beegle et al. (2012); Wollburg et al. (2020)
HH yearly per capita income	Cumulative yearly household per capita income (1,000 PPP$)	+	+	+	Meyer et al. (2018)

Source: Authors' own representation

4. Data

4.1 Study are and data collection

The survey instrument adheres to the design of a Living Standard Measurement Study (LSMS) survey (Grosh & Glewwe, 2000) extended by modules on shocks, risks and behavioural aspects pertaining to development. Aside from the usual asset, income and consumption modules, detailed household member characteristics as well as small-scale self-employment and wage

employment modules are included. Additionally, the survey instrument collects information on the financial state of the household by collecting data on borrowing, lending, savings, and public transfers.

The survey area consists of the provinces of Buriram, Ubon Ratchathani and Nakhon Phanom in Thailand and Ha Tinh, Thua Thien Hue and Dak Lak in Vietnam. These areas are characterized as being representative for rural Northeast Thailand and Central Vietnam (Hardeweg et al., 2013).

The basis for our analysis is the 2017 household survey, which consists of 10 sections, and covers 3,812 households. The printed version of the questionnaire consists of 81 pages with the computerized version containing over 900 individual variables. An average interview captured 1,524 data items due to a number of sections containing multiple rows of data (e.g. on each individual household member/occupation).

The computerized questionnaire included complex routing and over 450 plausibility rules. Warning messages appeared when such rules were violated and an overview of issues was highlighted on the final screen prior to interview synchronization. While interviews that violated plausibility rules could still be uploaded, interviewers were instructed to resolve any issues directly or enter a comment confirming the entry. The “no answer” code was implemented to record unwillingness to respond and was determined to only be used deliberately, albeit cautiously, in situation in which careful probing to elicit a response was not successful.

The survey was conducted in the field by teams of four to five interviewers operating under one experienced sub-team leader. In Thailand, there were a total of 47 interviewers allocated to ten teams: Four in each of the larger provinces of Buriram and Ubon Ratchathani and two in

Nakhon Phanom. In Vietnam, 53 interviewers were separated into 9 teams: three for each of the three provinces of Ha Tinh, Thua Thien Hue and Dak Lak. Data collection began in Thailand and lasted for five weeks whilst the survey in Vietnam lasted seven weeks and was offset by one week. After survey activities in Nakhon Phanom (Thailand) and Ha Tinh (Vietnam) had concluded, interviewer teams were relocated to Ubon Ratchathani and Thua Thien Hue and assisted the original interviewer teams during the final weeks of the survey. Interviews were uploaded to a separate database on a daily basis after being subjected to initial supervision instruments in the form of evening group discussions and manual reviews by interviewers. They were then reintroduced to the survey's data quality control process (e.g. in-depth reviews by data checking assistants) on the following day (see Figure A5). This study utilizes the data that was uploaded to a separate database from that of the main survey and we hence coin the interviews utilized in our study as raw interview data, which still contain the majority of non-sampling errors that occurred throughout each individual interview. By comparing raw interview data with the final data sets that were processed and cleaned, we are able to identify and categorize non-sampling errors in adherence to survey documentation and guidelines.

Additional paradata were generated throughout several stages of the survey (see Figure 1). First, during the interviewer training, paradata consisting of examinations of interviewers, in depth interviewer information and self-assessed interviewer personality traits were compiled. Second, during data collection, the interviewer and respondent individually evaluated the interview and the interaction with their counterpart. Third, after the conclusion of data collection, sub-team leaders evaluated interviewers based on their performance during training and data collection. The evaluation is based on their daily interaction with interviewers and

their presence as an observer in some 30% of interviews. Further, the sub-team leaders assessed the personality traits of each interviewer in their team, which could then be compared with the assessment on behalf of each interviewer. Additionally, the 2017 survey first provided a module on the personality traits of the respondent with personality trait items being based on the Big Five model developed by Costa and McCrae (1997).

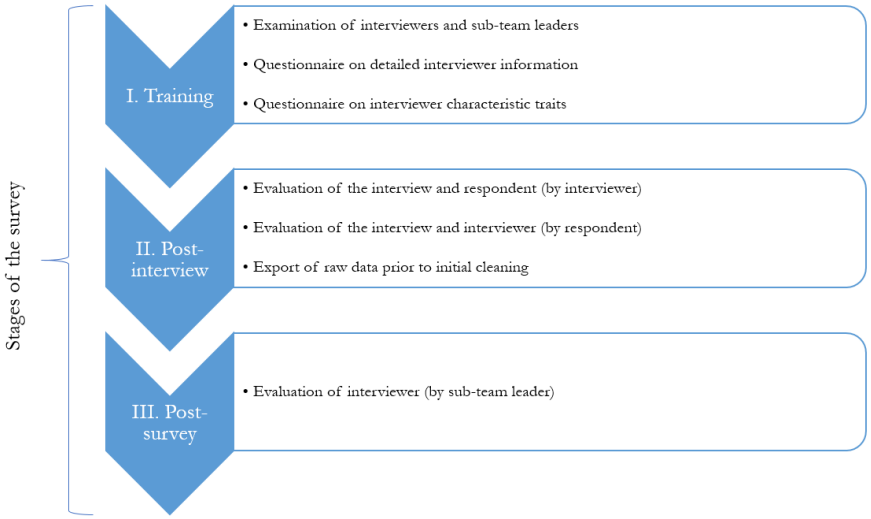

Figure 1. Overview of supplemental household survey paradata[2]

4.2 Data description

Table 2 presents the descriptive statistics of the variables used in our analysis. Regarding the dependent variables, we observe that the mean count of missing data per interview has dropped significantly from when compared with the earlier waves from 2007 and 2008. In spite of a higher average count of ~1,500 data items per interview in 2017, compared to the ~900 items in 2007 and 2008, we find that the mean count of missing items has dropped from 57 to 16 in Thailand and from 111 to 9 in Vietnam (Phung et al., 2015). Interviews were free of missing data in 140 cases (~8%) in Thailand and 38 cases (~2%) in Vietnam (see Figure A6). The average number of refusal cases is low in both countries, but significantly higher in Vietnam with approximately half of the interviews in Thailand and 30% in Vietnam being absent of

2 The questionnaires and materials used to collect the supplemental paradata are available on request.

refusals (see Figure A6). Conversely, implausible error counts were significantly lower in Vietnam with an interview containing on average 22 cases compared to the 28 observed in Thailand. The majority of non-sampling errors are found to stem from household and member characteristics as well as more complex survey modules pertaining to agriculture, income and consumption (see Figure A7). An overview of our dependent variables provided in Figure 2 shows that both missing data and implausible values peak in the initial weeks of the survey and decline as the survey progresses. Differences in the overall count of dependent variables between the two countries, in particular in the initial weeks, can be explained by the delayed start of the survey in Vietnam. Thus, allowing for lessons learned to be applied and reducing the count of missing data and implausible values. However, the opposite can be observed for refusal, which is shown to increase in later stages of the survey.

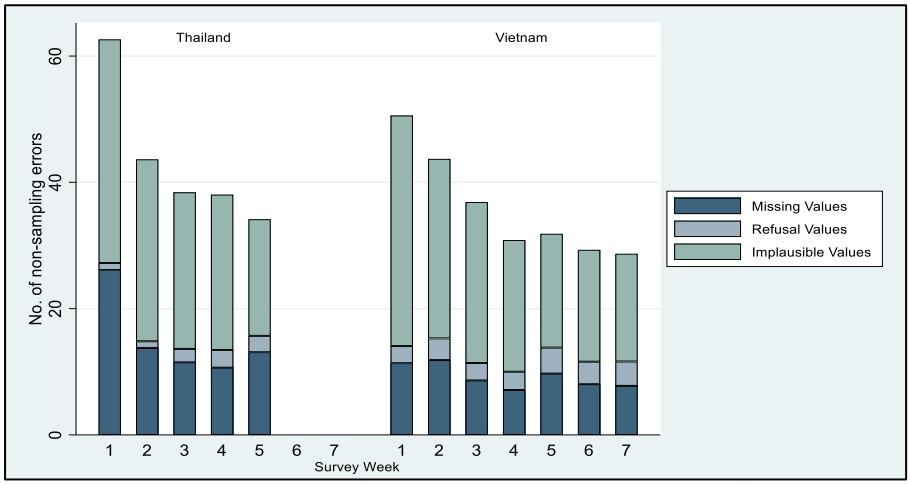

Figure 2. Mean non-sampling errors per interview and survey week

The majority of respondents are above the age of 50, with the average respondent being significantly older in Thailand. The ratio of female respondents is high in both countries (~60%) and respondents, on average, have 5 years of schooling in Thailand with Vietnamese respondents visiting school for on average an additional 2 years in comparison. Ethnic minority households are scarce in Thailand with 97% of respondents being of Thai heritage, whereas a significantly larger share of ethnic minorities (~21%) was present in the Vietnamese sample.

For example, ~9% of respondents that reside in the province of Dak Lak are of Ede heritage. The survey primarily targets household heads as respondents and was able to do so in 60% of interviews. In Vietnam, the majority of interviewed heads were male (70%), whereas the gender ratio of heads was almost even in Thailand. On average, respondents were interviewed 2-3 times in prior waves of the survey. Of those interviewed in 2016, 70% were reinterviewed in 2017 in Thailand and 54% in Vietnam. Personality traits show significant differences between the two countries with Thai respondents self-assessing themselves as being more open and exerting less neurotic and extraverted behaviours.

In both countries, interviewers can be characterized as being young with the majority of interviewers being below the age of 25, in particular in Thailand. In Thailand, interviewers are predominantly female, whereas the share of male interviewers is significantly higher (40%) in Vietnam. Interviewers are almost exclusively representatives of Thai or Kinh origin and are accordingly represent the ethnic majority group within the context of their country. Further, almost 50% are natives to the survey province in Vietnam with the share being significantly lower in Thailand (20%). Interviewers in Vietnam tend to be more experienced both in terms of education and previous experience as interviewers with on average three years of experience in other surveys. In Thailand, interviewers have at most 3 years of experience in the field. However, 10% of Thai interviewers were previously employed in previous waves of TVSEP, whereas interviewer continuity is significantly lower in Vietnam. Therefore, the cadre of interviewers from Thailand can be characterized as being younger, inexperienced students, while those in Vietnam have a more professional background with many interviewers being experienced, full-time interviewers. In terms of personality traits, interviewers in both countries can be characterized as being socially outgoing, cooperative, polite, curious and kind. There appear to be significant differences between the two countries with openness and agreeableness

on average being significantly higher for Vietnamese interviewers. The average interview duration was slightly under three hours in Thailand and significantly higher in Vietnam with an average of approximately four and a half hours. Furthermore, the number of survey items answered per minute is significantly higher for Thai interviewers. Their average number of entries per minute of ten significantly exceeds the lower count of seven in Vietnam. The difference in entry speed is at least partially driven by a higher share of interviews that were completed within one hour in Thailand (see Figure A8). Additionally, questionnaire complexity may explain the slower pace of Vietnamese interviews. For example, while the average aggregate size of agricultural land is significantly lower in Vietnam (~9,000m²) compared to Thailand (~23,000m²), the number of individual plots is higher with up to 16 plots of agricultural land being reported in Vietnam. Interviews were infrequently conducted in the presence of others and significantly less often in Vietnam (~10%) than Thailand (~20%). Significant technical malfunctions were experienced during 20% of interviews as the robustness of tablets was limited due to financial constraints of the survey. The average household in both countries consists of four to five members with yearly per capita income being comparable at approximately 3,000 PPP$ in both countries.

Table 2. Summary statistics

Variables	Thailand		Vietnam		
	Mean	Std. dev.	Mean	Std. dev.	T
Dependent Variables					
Missing data (per interview)	16.00	46.11	9.33	24.47	5.36***
Refusals (per interview)	1.87	6.17	3.42	4.24	-8.66***
Implausible values (per interview)	28.39	19.26	22.31	17.36	9.76***
Independent Variables					
Respondent					
Age (years)	57.86	12.76	52.99	13.87	11.05***
Gender (1=male, 0=female)	0.34	0.48	0.43	0.50	-5.68***
Years of education	5.42	2.96	6.75	3.75	-11.85***
Ethnicity (1=Thai|Kinh, 0=other)	0.97	0.18	0.79	0.41	17.18***
Head of Household (1=yes, 0=no)	0.57	0.49	0.57	0.49	0.00
Number of times interviewed	2.51	0.79	2.41	0.74	3.66***
Openness (scale 1-7)	4.61	1.27	4.05	1.38	12.63***
Extraversion (scale 1-7)	4.49	1.05	4.56	1.10	-1.87*
Neuroticism (scale 1-7)	3.32	1.12	4.42	1.07	-30.33***
Interviewer					
Age (years)	22.31	1.99	24.88	2.33	-35.80***
Gender (1=male, 0=female)	0.28	0.45	0.40	0.49	-7.68***

Variables	Thailand		Vietnam		
	Mean	Std. dev.	Mean	Std. dev.	T
Years of education	15.31	1.19	16.12	1.26	-20.01***
Ethnicity (1=Thai|Kinh, 0=other)	0.96	0.20	0.97	0.16	-2.64***
Survey experience – Other (1=yes, 0=no)	0.40	0.49	0.88	0.32	-34.93***
Survey experience – TVSEP (1=yes, 0=no)	0.13	0.33	0.04	0.21	8.93***
Years of survey experience	0.79	0.93	2.91	2.43	-34.79***
Born locally (1=yes, 0=no)	0.45	0.50	0.21	0.41	15.99***
Training performance evaluation (scale 1-7)	6.12	0.64	6.19	0.59	-3.17***
Openness (scale 1-7)	4.42	0.67	4.47	0.58	-2.54**
Extraversion (scale 1-7)	3.80	0.48	3.81	0.37	-0.74
Agreeableness (scale 1-7)	5.15	0.64	5.83	0.59	-33.34***
Field of study					
Agriculture|Economics (1=yes, 0=no)	0.22	0.41	0.52	0.50	-19.48***
Sociology|Languages|Education (1=yes, 0=no)	0.39	0.49	0.43	0.50	-2.79***
Administration|Politics|Law (1=yes, 0=no)	0.39	0.49	0.05	0.22	27.30***
Interview/Survey					
Interview duration (minutes)	165.05	56.75	274.59	95.99	-41.89***
Answers per minute	10.06	3.56	6.95	2.09	32.16***
Morning interview (1=yes, 0=no)	0.53	0.50	0.59	0.49	-3.95***
Others present during interview (1=yes, 0=no)	0.21	0.41	0.12	0.33	7.27***
Very negative impact of tablet malfunction (1=yes, 0=no)	0.23	0.42	0.18	0.39	3.33***
Survey week	2.65	1.14	4.06	1.51	-31.91***
Household size (no. of members)	4.57	1.91	4.49	1.79	1.33
Household agricultural land size (1,000m²)	22.83	26.45	9.23	29.95	14.53***
Yearly per capita income (1,000 PPP$)	3,214.76	6,045.62	2,936.44	4,316.58	1.60
Provinces (Thailand|Vietnam):					
Buriram|Ha Tinh (1=yes, 0=no)	0.47	0.50	0.34	0.48	-
Ubon Ratchathani|Thua Thien Hue (1=yes, 0=no)	0.37	0.48	0.31	0.46	-
Nakhon Phanom|Dak Lak (1=yes, 0=no)	0.17	0.37	0.35	0.48	-

Notes: * Significant at 10%.; ** Significant at 5%.; *** Significant at 1%.
1,816 observations in Thailand/1,830 observations in Vietnam.
Source: Authors' calculations based on TVSEP survey 2017 (https://www.tvsep.de)

5. Results

Table 3 reports the determinants of non-sampling errors using the combined sample of Thailand and Vietnam. The results of the models on missing data, refusals and implausible values are summarized as follows:

The results of the first model on missing data show that respondent characteristics such as age, gender, education and ethnicity do not significantly affect their expected count, which is in line with our hypothesis. In contrast, the majority of interviewer characteristics are significant determinants of missing data. Increasing interviewer age is found to coincide with a higher expected count of missing data, which may be explained by interviewers above the age of 25 assessing themselves as being inexperienced in the use of tablets in the complementary survey paradata. Further, we observe a significant effect of increasing interviewer openness, which

matches our expectations. Being open-minded is associated with higher levels of creativity, but simultaneously also with a lack of focus. Conversely, a lower score indicates a more analytical and focused approach in an individual's behaviours. Interviewer experience in previous waves of TVSEP is demonstrated to significantly reduce the expected count of missing data, which is consistent with the general consensus in the literature. Prior experience in the use of CAPI and the survey instrument facilitates the completion of high-quality interviews. By controlling for respondent-interviewer dyads and congruency of gender, we find that interviews with female respondents and male interviewers result in a significantly higher expected count of missing data. This goes against our hypothesis and Phung et al. (2015), who find that male interviewers generally outperformed their female counterparts in earlier waves of TVSEP. The ethnicity of interviewers negatively and highly significantly affects the expected count of missing data with interviewers of Thai or Kinh descent outperforming those of minority heritage. While this is in line with our hypothesis, we argue that further examination at the country-level is required in order to determine whether this effect is driven by ethnicity itself or by interactions of respondent and interviewer heritage. Accounting for such an interaction is only possible at the country-level for the Vietnamese sample due to a very small count of some 50 minority respondents in the Thai sample (~3%). Interviewers with an educational background in economics and agriculture, which matches the subject of the survey, are shown to conduct interviews with a significantly lower expected count of missing data. In terms of the interview environment, faster entry speeds are shown to coincide with a higher count of implausibly skipped items, which matches our hypothesis. Interviewers with above average entry speeds either act carelessly during data entry or are willing to accept impairments in terms of interview quality (e.g. in order to reduce their workload), as argued by Olson & Peytchev (2007). Furthermore, the timing of the interview is shown to significantly influence the expected count of missing data with morning interviews being of higher quality than afternoon or evening

sessions, which mirrors the findings of Phung et al. (2015). The progression of the survey, as proxied by means of the survey week, is shown to be highly significant. As the survey progresses, we argue that the knowledge and familiarity of the interviewer in terms of the computerized questionnaire increases, which facilitates the completion of full interviews with minimal to no entry errors (Townsend et al., 2013). Finally, of the three proxies included that capture questionnaire complexity, the overall size of the household is shown to significantly increase the count of missing data. This matches our hypothesis and can be explained by the multi-layered sub-sections on household members and the multitude of items contained therein.

Regarding the determinants of refusals, we find that older respondents are more likely to cooperate and provide full answers than their younger counterparts However, this effect is shown to decline in older cohorts. Moreover, we observe that respondents who have at least completed secondary education have a lower expected count of refusals. Against expectations, heads of household are found to be less likely to cooperate during interviews. Additionally, personality traits are found to affect cooperation during interviews. For example, respondents who possess neurotic traits such as being easy to frustrate, anxious or impatient, are found to result in less cooperative interviews. Conversely, extraverted respondents, who are characterized as socially outgoing and communicative are not found to be more cooperative than introverted respondents. The personality of interviewers is found to be similarly important with more sympathetic and cooperative persons significantly improving the cooperation of respondents than those that scored lowly on the scale of agreeableness. Interviewer field of study is again found to be a highly significant determinant of refusal. However, interviewers with a background in sociology, languages and entertainment are found to outperform their peers in eliciting responses from respondents when compared with their peers. Interviewers of Thai or Kinh descent provide interviews with a lower expected count of refusals and while the

literature suggests that congruency, in particular of gender, can improve cooperation and reliability of data provided (e.g. Catania et al., 1996), we are unable to confirm this in our model on refusals. We argue that congruency of traits and characteristics likely plays a greater role in surveys in which more sensitive topics are included in the survey instrument, which may explain the deviation from the literature. In controlling for the interview environment, we find that that longer interview times signal a lower expected count of refusal while the speed of entry does not have a significant effect. We find a statistically significant effect of the survey week, which suggests that interviewers encountered fatigue at later stages of the survey, manifesting in lower response rates. In line with the findings of literature on sensitivity of income data in surveys, we find that households that are more well-off in terms of per capita income have a higher expected count of refusal. Meyer et al. (2018), for example find that households that are more well-off are less likely to disclose full information on sources of income. This is also reflected in the data with the highest average count of refusals per interview being found in modules on household finance.

The third model on implausible values finds that household heads provide interviews with a higher expected count of measurement error, which does not match our expectations. While Phung et al. (2015) found that household heads do not significantly reduce nor increase the prevalence of non-sampling error, we argue that the 9-year gap between the survey waves, which goes hand in hand with an aging population of heads by some 10 years, may explain the change of significance of this variable. The average head being significantly older than the average non-head respondent and having fewer years of education in 2017 survey, which would be in line with the argumentation that decreased cognitive ability results in less reliable and accurate data (Knäuper et al., 1997; Krosnick, 1991). Respondent continuity, as proxied by the number of times a respondent has been interviewed, is found to significantly reduces the

expected count of implausible values. Interviewers and respondents who are characterised as being open are shown to significantly increase the expected count of implausible values, which matches the results of the missing data model. In addition, extraverted interviewers were able to collect more reliable data. Interviewer age and experience were expected to significantly improve the quality of data and while this can be confirmed for age, interviewer experience in other surveys is found to lead to a higher expected count of implausible values. While being against the general consensus of the literature, this is in line with Fowler and Mangione (1990), who argue that inexperienced interviewers often adhere more closely to existing survey guidelines resulting in higher quality data. Higher ranking in terms of the ex-ante evaluation of interviewer performance during training, which consisted of indicators such as punctuality, active participation and understanding of the survey instrument, was shown to significantly reduce the expected count of implausible values. As in the first model, local interviewers are observed to collect data with a higher expected count of error. Additionally, interviewers with an educational background that matches the topic of the survey and ethnic Thai and Kinh produce higher quality interviews. Longer interviews resulted in a higher expected count of implausible values – likely as a direct result of respondent and interviewer fatigue. Conversely, faster entry times are associated with less accurate responses. As hypothesized, the presence of others during an interview leads to an increasing prevalence of implausible data, which is likely the result of a lack of confidentiality during the interview (Krumpal, 2013). The progression of survey activities is shown to lead to a significant reduction in the expected count of implausible values, which matches the results of our first model. Once more, individuals/households who have higher incomes are shown to be less likely to report accurately on their income, as found by Meyer et al. (2018).

Across all three model variants, we observe statistically significant differences in the country indicator variable, which, as argued in section 3, warrants further examination. Furthermore, accounting for differences between the two countries in terms of interviewer characteristics, survey environment and culture (see Table 2) are expected to provide a more comprehensive overview of determinants of non-sampling error.

Table 3. Combined sample negative binomial regression results (N=3,633)

	(1) Missing data	(2) Refusals	(3) Implausible values
Respondent			
Age (years)	-0.007	-0.048***	0.000
	(0.009)	(0.011)	(0.004)
Age squared	0.000	0.000***	0.000
	(0.000)	(0.000)	(0.000)
Gender (1=male, 0=female)	0.027	-0.208**	-0.019
	(0.049)	(0.069)	(0.025)
Secondary education and above (1=yes, 0=no)	0.015	-0.188**	0.037
	(0.049)	(0.061)	(0.023)
Ethnicity (1=Thai|Kinh, 0=other)	-0.028	-0.548***	0.054
	(0.065)	(0.080)	(0.031)
Head of household (1=yes, 0=no)		0.165*	0.068**
		(0.066)	(0.024)
Number of times interviewed		-0.026	-0.019***
		(0.015)	(0.006)
Openness (scale 1-7)			0.012
			(0.007)
Extraversion (scale 1-7)		-0.040	
		(0.022)	
Neuroticism (scale 1-7)		0.055*	
		(0.023)	
Interviewer			
Age (years)	0.057***	0.001	-0.028***
	(0.009)	(0.012)	(0.004)
Gender (1=male, 0=female)	0.138**	-0.204**	0.028
	(0.054)	(0.067)	(0.025)
Years of education	-0.051**	-0.018	-0.006
	(0.017)	(0.021)	(0.008)
Ethnicity (1=Thai|Kinh, 0=other)	-0.594***	-0.746***	-0.336***
	(0.110)	(0.137)	(0.052)
Born locally (1=yes, 0=no)	0.147**	-0.037	0.186***
	(0.045)	(0.059)	(0.022)
Training performance evaluation (scale 1-7)		-0.025	-0.089***
		(0.043)	(0.016)
Field of study (Social field is base)			
Agriculture/Economics	-0.133**	0.333***	-0.081***
Politics/Administration/Law	(0.051)	(0.060)	(0.024)
	-0.165**	0.400***	-0.109***
	(0.054)	(0.070)	(0.026)
Openness (scale 1-7)	0.182***		0.072***
	(0.032)		(0.016)
Extraversion (scale 1-7)			-0.124***
			(0.022)
Agreeableness (scale 1-7)		-0.358***	-0.019
		(0.042)	(0.017)

Survey experience – Other (1=yes, 0=no)#Years of survey experience	-0.007	0.018	0.012*
	(0.011)	(0.017)	(0.006)
Survey experience – TVSEP (1=yes, 0=no)#Years of survey experience	-0.172***	0.027	-0.029
	(0.032)	(0.043)	(0.015)
Congruent characteristics			
Respondent gender#Interviewer gender (male/male)	0.064	0.188	-0.073
	(0.080)	(0.103)	(0.038)
Interview/Survey			
Interview duration (minutes)	0.000	-0.001**	0.001***
	(0.000)	(0.000)	(0.000)
Answers per minute	0.017*	-0.002	0.021***
	(0.008)	(0.011)	(0.004)
Morning interview (1=yes, 0=no)	-0.114**	0.082	0.027
	(0.038)	(0.048)	(0.018)
Others present during interview (1=yes, 0=no)	-0.041	-0.074	0.061*
	(0.052)	(0.066)	(0.024)
Very negative impact of tablet malfunction (1=yes, 0=no)	0.073	0.033	-0.018
	(0.049)	(0.061)	(0.023)
Survey week	-0.150***	0.107***	-0.129***
	(0.014)	(0.018)	(0.007)
Country (1=Vietnam, 0=Thailand)	-0.417***	0.684***	-0.176***
	(0.081)	(0.104)	(0.041)
Household characteristics			
Household size (no. of members)	0.040***	-0.019	-0.013*
	(0.012)	(0.015)	(0.006)
Household agricultural land size (1,000m²)	0.001	0.001	-0.000
	(0.001)	(0.001)	(0.000)
Yearly per capita income (1,000 PPP$)	0.002	0.011*	0.006***
	(0.002)	(0.005)	(0.002)
Constant	-5.076***	-2.085***	-2.055***
	(0.430)	(0.632)	(0.258)
/lnalpha	0.123	0.394	-1.484
	(0.024)	(0.035)	(0.027)
AIC	24632	14734	27790

* Significant at 10%; ** Significant at 5%; *** Significant at 1%.
Notes: Standard errors in parentheses

The results of the country-level analysis are presented in Table 4. In terms of respondent characteristics, the results of the two model variants on missing data mirror those of the combined model. Respondent characteristics are shown to be widely insignificant predictors of the expected count of missing data. At the country-level, the significance of interviewer age and education is shown to be driven solely by interviewers in Vietnam. We argue that the homogeneity of age that is inherent in the Thai sample of interviewers compared to the more diverse group of interviewers in Vietnam likely explains the differences in significance. Meanwhile, local interviewers are shown to yield a higher expected count of missing data, albeit only being significant predictors in Vietnam. Further, the effects of interviewer gender

are found to be inconsistent between the two countries: Interviews with female respondents are found to lead to higher quality data in dyads in which the interviewer is also female in Thailand. Whereas other combinations of gender are not found to significantly change the expected count of missing data. In addition, we observe that congruency of respondent and interviewer ethnicity plays a key role, albeit mainly in the Vietnamese model. Interviews in which both parties are of Kinh heritage are less afflicted with missing data than mismatched or congruent minority interviews, which is in line with our expectations. In both countries, prior experience in TVSEP is found to significantly reduce the expected count of missing data, which is consistent with the results of the combined model. Having an educational background in economics or agriculture was found to significantly reduce the expected count of missing data only for the Thai sample, in which interviewers consisted of younger, less experienced students. Conversely, the field of study plays a less significant role for the professionalised and more experienced Vietnamese interviewers. In terms of interview characteristics, the speed of entry loses significance at the country-level, while being a significant predictor at the combined model. Taking into consideration that the combined model pools together the experienced Vietnamese interviewers and Thai students, we can assume that the significance in the combined model is driven by faster speeds of entry on behalf of less experienced interviewers resulting in a higher expected count of missing data when compared with those that are experienced. Finally, the complexity of the interview is shown to influence the expected count of missing data in both countries.

In terms of refusals, the direction of determinants remains consistent in the country-level model variants. However, at the country-level household heads are no longer expected to be more likely to refuse to provide an answer. In addition, the results show that extraverted respondents have a reduced expected count of refusal cases, whereas the significance of neurotic traits is no

longer significant. Interviewer characteristics such as having a background in a social field of study and personality traits such agreeableness are shown to consistently reduce the expected count of refusals at the country-level. In contrast, other determinants are shown to vary between the two countries. While older interviewers are more prone to encounter refusal in Thailand, the opposite can be observed in the Vietnamese sample, which likely coincides with the differences in the pools of interviewers hired. In the culturally diverse country of Vietnam, local interviewers are more likely to gain the trust and levels of cooperation necessary to illicit a response, whereas local Thai interviewers have a higher expected count of refusals. Survey experience, albeit being limited in scope, is found to significantly reduce cooperation in Thailand, whereas the more experienced interviewers in Vietnam are more likely to elicit a positive response. Furthermore, gender effects are found to be significant in Vietnam with interviews between males leading to a higher count of expected refusals and interviews between female interviewers and male respondents leading to a lower count of refusals when compared to the reference of interviews between females. In addition, in Thailand, interviewer fatigue is seemingly a prevalent issue with the count of refusals increasing as the survey progresses whereas the opposite is the case in Vietnam. It is likely that Vietnamese interviewers are more used to the conditions in the field during prolonged periods of data collection and are thus able to consistently illicit responses from Vietnamese respondents. Against expectations, ethnicity does not seem to be a significant determinant of refusal in Vietnam, unlike in the models on missing data and implausible values. As the average count of refusals is exceptionally low in both country-level samples, it appears that cultural differences in the context of the Vietnamese survey do not significantly increase non-cooperation in the form of refusals, but rather reduce the quality of responses.

The country-level model of implausible values is for the greater part consistent with the combined model. In terms of respondent and interviewer characteristics, we reaffirm that heads of household, local interviewers and interviewers characterised as lacking focus are significant predictors of the expected count of implausible values. In Thailand, respondent continuity, good performance of interviewers during training, having studied or studying economics or agriculture, and extraverted behaviours are found to significantly reduce the number of implausible values. In addition, student interviewers in Thailand performed better, the more advanced their studies were and the more sympathetic and cooperative their personality was. Regarding the professional cadre of interviewers in Vietnam, we find that determinants of reliable and accurate data differ from those of the Thai interviewers. Firstly, Vietnamese interviewers perform significantly better if they have an educational background in fields such as sociology, languages or education. Secondly, older interviewers are associated with a significantly lower expected count of implausible values, as are those with prior experience in the TVSEP survey. Performing well in post-survey training schedules is found to be a consistent determinant of higher quality interviews. Congruency of ethnicity is found to be of high importance in the Vietnamese sample. Interviews between Kinh respondents and minority interviewers result in a significantly higher count of implausible values. Additionally, interviews with congruent Kinh interviewer-respondent dyads are of significantly higher quality than minority-minority dyads. This is in line with our hypothesis and matches the findings of Adida et al. (2016), Feskens et al. (2017) and Pennell et al. (2017) who find that the social norms of countries with diverse ethnic groups can affect interactions between respondents and interviewers that result in a reduced level of cooperation with individuals from outside of their community. However, the ethnic background of the interviewer does not seem to play a significant role in interviewing minority households. Interviews with female respondents and male interviewers are found to be more prone to unreliable or inaccurate

responses than matching female dyads, which further solidifies that congruency of characteristics plays a vital role in reducing the prevalence of non-sampling errors. Personality traits, while being highly significant in the Thai sample are only relevant in terms of agreeableness in Vietnam and the result that more sympathetic and cooperative interviewers provide interviews of lesser quality does not match our expectations. A potential explanation lies in the distribution of agreeableness, which is highly skewed towards the upper side of the scale in Vietnam. Prior experience in other surveys is found to have a negative effect on the quality of data in Vietnam, which following the findings of Fowler and Mangione (1990) can be explained by experienced interviewers applying survey procedures and guidelines of other surveys rather than conforming to those that they are currently employed in. The interview and survey environment are also shown to be a significant predictor of implausible values at the country-level. Longer interviews as well as an increased frequency of data entry, which simultaneously suggests that respondents are provided with less time to construct their response, are shown to result in a higher expected count of implausible values. Additionally, the presence of others, while only being significant in the Vietnamese sample, is found to lead to a higher count of error. In the Thai sample, more complex income structures of the household as proxied for by yearly per capita income are shown to increase the expected count of implausible values, whereas against expectations, increasingly larger households are found to result in higher-quality interviews in Vietnam. Furthermore, tablet malfunctions that interviewers assess as being very negative are found to reduce the expected count of implausible values in the Vietnamese sample, which may be explained by more professional interviewers being able to handle situations in which data are lost and being more careful in completing interviews in which technical malfunctions occurred.

In examining the provincial indicators one can clearly observe that there are significant differences in terms of provinces. We argue that the significance of the provincial dummies can be explained as follows: In terms of survey management, the provincial team leaders in Buriram and Ubon Ratchathani had extensive experience in their roles in previous waves of TVSEP, whereas in Nakhon Phanom a younger, less experienced manager was hired. In Vietnam, we argue that the significance of the provincial dummies can be explained in part by the survey schedule. Initially, survey activities began in Ha Tinh before interviewer teams moved on to Dak Lak and finally concluded the data collection process in Thua Thien Hue. Accordingly, Ha Tinh has a higher count of missing data and implausible values, whereas the provinces scheduled later on during the survey are shown have a significantly higher count of refusals. Further, differences in the complexity of agricultural activities are likely drivers of non-sampling errors in Vietnam. For example, the mean number of unique crops planted per household is eleven in Ha Tinh. In comparison, the provinces of Thua Thien Hue and Dak Lak are less complex with on average nine and five unique crops planted per household. Furthermore, land parcels in Dak Lak are less fragmented with the average household having two plots of ~3,800m² whereas the other provinces have on average three to four plots of ~ 1,500m², which increases the complexity of the response process on behalf of the respondent.

Table 4. Country-level negative binomial regression results (N=1,806|1,827)

	Thailand			Vietnam		
	(1) Missing data	(2) Refusals	(3) Implausible values	(1) Missing data	(2) Refusals	(3) Implausible values
Respondent						
Age (years)	-0.005	-0.028	-0.007	-0.008	-0.042***	0.008
	(0.018)	(0.020)	(0.007)	(0.010)	(0.013)	(0.006)
Age squared	0.000	0.000	0.000	0.000	0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gender (1=male, 0=female)	-0.016	-0.063	0.011	0.014	-0.358***	-0.056
	(0.078)	(0.105)	(0.032)	(0.056)	(0.087)	(0.039)
Secondary education and above	-0.053	0.063	-0.006	0.135**	-0.235***	0.051
(1=yes, 0=no)	(0.098)	(0.123)	(0.037)	(0.051)	(0.066)	(0.029)
Ethnicity (1=Thai|Kinh, 0=other)	0.318	0.267	0.049	0.459	0.142	0.632***
	(0.190)	(0.259)	(0.074)	(0.300)	(0.373)	(0.185)
Head of household (1=yes, 0=no)		0.097	0.061*		0.082	0.082*
		(0.095)	(0.030)		(0.088)	(0.039)
Number of times interviewed		-0.023	-0.020**		-0.015	-0.020*
		(0.022)	(0.007)		(0.021)	(0.009)
Openness (scale 1-7)			-0.000			0.013
			(0.010)			(0.010)
Extraversion (scale 1-7)		-0.080*			-0.034	
		(0.035)			(0.025)	
Neuroticism (scale 1-7)		0.033			0.029	
		(0.035)			(0.028)	
Interviewer						
Age (years)	0.029	0.098***	0.011	0.026**	-0.040**	-0.047***
	(0.022)	(0.025)	(0.008)	(0.010)	(0.014)	(0.006)
Gender (1=male, 0=female)	0.264**	-0.015	-0.067	-0.019	-0.589***	0.126***
	(0.096)	(0.110)	(0.038)	(0.064)	(0.090)	(0.038)
Years of education	-0.058	-0.056	-0.046**	-0.052**	0.036	-0.009
	(0.038)	(0.044)	(0.015)	(0.018)	(0.024)	(0.010)
Ethnicity (1=Thai|Kinh, 0=other)	-0.292	-0.693**	-0.232***	-0.260	-0.558	0.325
	(0.178)	(0.217)	(0.069)	(0.271)	(0.335)	(0.170)
Born locally (1=yes, 0=no)	0.110	0.194*	0.202***	0.294***	-0.248**	0.019
	(0.079)	(0.097)	(0.030)	(0.065)	(0.083)	(0.037)
Training performance evaluation		-0.080	-0.070***		-0.045	-0.188***
(scale 1-7)		(0.069)	(0.021)		(0.053)	(0.024)
Field of study (Social field is base)						
Agriculture/Economics	-0.435***	0.419***	-0.255***	0.002	0.282***	0.070*
	(0.096)	(0.114)	(0.037)	(0.058)	(0.077)	(0.033)
Politics/Administration/Law	-0.276***	0.201*	-0.174***	-0.121	-0.100	0.204**
	(0.076)	(0.100)	(0.031)	(0.118)	(0.162)	(0.073)
Openness (scale 1-7)	0.058		0.087***	0.070		0.040
	(0.063)		(0.025)	(0.045)		(0.028)
Extraversion (scale 1-7)			-0.116***			-0.030
			(0.030)			(0.039)
Agreeableness (scale 1-7)		-0.251**	-0.079**		-0.254***	0.061*
		(0.084)	(0.025)		(0.053)	(0.025)
Survey experience – Other	-0.005	0.357***	-0.006	0.014	-0.050**	0.021**
(1=yes, 0=no)	(0.043)	(0.050)	(0.017)	(0.011)	(0.015)	(0.007)
# Years of survey experience						
Survey experience – TVSEP	-0.164**	0.106	0.012	-0.160***	-0.118	-0.130***
(1=yes, 0=no)	(0.050)	(0.063)	(0.020)	(0.046)	(0.063)	(0.027)
# Years of survey experience						
Congruent characteristics						
Respondent gender	0.173	-0.100	-0.046	0.148	0.418***	-0.062
# Interviewer gender (male/male)	(0.143)	(0.175)	(0.055)	(0.087)	(0.117)	(0.050)
Respondent ethnicity				-0.629*	-0.598	-0.714***
# Interviewer ethnicity (maj./maj.)				(0.302)	(0.377)	(0.186)
Interview/Survey						
Interview duration (minutes)	0.001	-0.000	0.001*	0.000	-0.002***	0.001***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Answers per minute	0.016	0.030	0.012*	-0.002	-0.085***	0.034***
	(0.014)	(0.019)	(0.006)	(0.016)	(0.020)	(0.010)
Morning interview (1=yes, 0=no)	-0.113	0.127	0.022	-0.131**	0.042	0.012
	(0.062)	(0.074)	(0.024)	(0.043)	(0.057)	(0.025)
Others present during interview	-0.033	-0.064	0.020	-0.073	0.029	0.097*

(1=yes, 0=no)	(0.077)	(0.097)	(0.030)	(0.066)	(0.088)	(0.038)
Very negative impact of tablet	0.128	0.052	0.022	-0.127*	-0.082	-0.104**
malfunction (1=yes, 0=no)	(0.080)	(0.098)	(0.031)	(0.059)	(0.079)	(0.033)
Survey week	-0.241***	0.271***	-0.159***	-0.141***	-0.056*	-0.096***
	(0.028)	(0.035)	(0.011)	(0.021)	(0.028)	(0.012)
Provinces (Thailand|Vietnam):						
Ubon Ratchathani|Thua Thien Hue	0.111	0.137	-0.008	0.302***	0.729***	-0.025
(ref: Buriram|Ha Tinh)	(0.091)	(0.102)	(0.035)	(0.081)	(0.107)	(0.047)
Nakhon Phanom|Dak Lak	0.405***	-0.185	-0.063	-0.311***	0.795***	-0.251***
(ref: Buriram|Ha Tinh)	(0.119)	(0.154)	(0.050)	(0.072)	(0.094)	(0.042)
Household characteristics						
Household size (no. of members)	0.047*	-0.027	-0.009	0.007	-0.010	-0.021*
	(0.020)	(0.025)	(0.008)	(0.015)	(0.020)	(0.009)
Household agricultural land size	-0.000	-0.001	-0.000	0.001	0.001	-0.000
(1,000m²)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Yearly per capita income (1,000 PPP$)	0.001	0.015*	0.007***	0.019***	-0.013	0.005
	(0.003)	(0.006)	(0.002)	(0.005)	(0.007)	(0.003)
Constant	-4.486***	-6.319***	-1.591***	-4.256***	-0.743	-2.773***
	(0.846)	(1.130)	(0.039)	(0.589)	(0.934)	(0.466)
/lnalpha	0.409	0.505	-1.591	-0.440	-0.003	-1.572
	(0.033)	(0.056)	(0.039)	(0.037)	(0.050)	(0.041)
AIC	12,707	6,090	14,125	11,474	8,297	13,427

* Significant at 10%; ** Significant at 5%; *** Significant at 1%.
Notes: Standard errors in parentheses

The results of our combined and country-level regression models demonstrate that determinants from all four specified factors (interviewer and respondent characteristics; interview environment; and survey environment) significantly affect non-sampling errors (see Table 3/4). Furthermore, the country-level models show that there are significant differences between determinants of non-sampling errors although an identical survey instrument was implemented.

In order to determine the relative importance of determinants in our models on non-sampling error we apply a model transformation developed by Long & Freese (2014), which transforms coefficients into percent change coefficients (see Figure A9-A11). The results indicate that prior experience in waves of the TVSEP are associated with a ~15% decrease of expected counts of missing data. Further, the count of missing data is expected to decline by 13-21% for each additional week of survey activities. Further, we find differing significance and relative importance of varying interviewer characteristics such as field of study (-30%), years of education (-5%), gender (30%), or being a member of a majority ethnic group (-25%).

Refusals are mainly driven by interviewer and respondent characteristics with ethnicity playing a key role in eliciting cooperation in interviews. Ethnic majority interviewers were able to

reduce the expected count of refusal by more than 40% and gender was determined to play a significant role in Vietnam, where male interviewers were able to increase the likelihood of cooperation by 30%. Additionally, interviewers without an educational background in the more social fields of study were less likely to elicit responses with counts of refusals increasing by 22-50%. The personality trait agreeableness was found to be highly significant in obtaining responses in interviews and more sympathetic interviewers reduced the expected count of refusals by 22%, which is consistent in both countries. In terms of respondent characteristics, an increase of respondent age by one year was found to result in a 3-4% decrease in the expected counts of refusals. Additionally, interviews with male (-45%) as well as highly educated (-21%) respondents resulted in significantly fewer refusals in Vietnam. The determinants of implausible values are found to be more consistent at the country-level in comparison to other types of non-sampling error. Interviews with heads of their respective household are found to yield a 6-9% increase in expected counts of implausible values and respondent continuity is found to significantly, albeit only slightly, decrease the prevalence of unreliable data. Interviewer characteristics are shown to be key determinants of implausible values. In particular, interviewer personality traits reflecting extraversion and agreeableness are found to significantly reduce the count of implausible data alongside high scores in terms of their performance during training being key, in particular in the Vietnamese sample (-17%). Moreover, ethnicity is found to play a key role in both countries with respondent-interviewer dyads in Vietnam, for example, leading to an expected increase of 88% in terms of implausible values when an ethnic Kinh respondent was interviewed by an incongruent minority interviewer. As reflected in the models on missing data, the expected count of implausible values decreases with each additional survey week by up to 15% and increasing interview length and speed of entry are found to result in more implausible values. Across all types of non-sampling error, significant provincial effects were observed. For example, missing data

were found to be more prevalent in interviews that took place in the provinces of Nakhon Phanom (50%) and Thua Thien Hue (35%). Conversely, interviews in Dak Lak contained ~30% fewer missing data than in the other provinces of Vietnam. Cases of refusal were subject to strong provincial effects in Vietnam with the likelihood of refusal being twice as high in interviews conducted in the provinces of Thua Thien Hue and Dak Lak when compared to the initial survey activities in Ha Tinh. In the final country-level model variant provincial effects are observed only in the province of Dak Lak in Vietnam with interviews containing 22% fewer implausible values.

6. Conclusion

In this paper, we addressed the impact of interviewer and respondent characteristics as well as characteristics of the interview and survey environment on non-sampling errors. We provide valuable insights regarding factors influencing three types of non-sampling error, namely missing data, refusals and implausible values.

Our results are able to validate the importance of interviewer and respondent characteristics. Experience of interviewers both accumulated prior to and during survey activities is shown to be a key determinant of non-sampling error. While experience in other surveys is shown to yield mixed effects, continuity of interviewers within the same survey is shown to significantly reduce the expected count of non-sampling error. Furthermore, we are able to provide novel insights on the impact of personality traits, in particular of interviewers. For example, extraversion and agreeableness are found to be important traits in interviewers with higher scores on the scale of these personality traits being associated with higher levels of cooperation and the collection of more reliable and accurate data. An educational background of interviewers that matches the subject of the survey is found to be of great import, particularly

for interviewers hired from a pool of young and less experienced students. Conversely, well-established, professional interviewers do not seem to profit significantly from a matching field of study. Although interviewers were randomly allocated to respondents, homogeneity of interviewer characteristics represented a bottleneck to our analysis. Nonetheless, we were able to establish a significant positive effect of congruent characteristics such as ethnicity, with mismatched dyads being a key source of non-sampling error in Vietnam. An experimental approach of interviewer-respondent allocations with a more diverse pool of interviewers in ensuing waves may facilitate future research on the influence of congruency of characteristics. While the interviewer is seemingly found to play a greater role in determining quality of interview data, interviews with household heads in the survey analysed in this paper are found to yield interviews of lower quality. The common approach of interviewing household heads in developing countries is generally expected to increase data quality in the literature due to their unique status within the household. We argue that the quality of response of household heads in long-term panels is likely to decrease as they age, which suggests that their quality as an "ideal" proxy respondent reaches a turning point throughout later waves. This warrants further attention into determining how to select an "ideal" respondent in a long-term panel and when contingencies should be made to target new respondents.

Additionally, we find that the introduction of a computerized questionnaire has drastically reduced the threat of missing data as found in Phung et al. (2015) and increased the overall quality of interviews. However, implausible values are identified as remaining a persistent threat in CAPI interviews that must be addressed by survey providers and designers. We argue, that while the implementation of plausibility rules and an intensification of supervision in the more recent TVSEP waves has significantly reduced non-sampling errors, data quality may

further be improved by extending and optimizing automated plausibility rules in order to provide data that is of the utmost quality to be used by researchers and policy-makers.

In conclusion, we advise that proxy respondents in household surveys in developing countries be selected according to their status in the household and their knowledge of household activities. Interviewers should mainly be selected based on their prior experience in survey work and continuity in the hiring process is recommended. Further, hired interviewers should exhibit traits of being outgoing, sociable and considerate. The importance of congruent characteristics must be considered, in particular in countries with diverse cultures. The interview and survey environment, while neglected in previous studies, significantly affects the prevalence of non-sampling error. For example, the entry speed of interviewers in CAPI is easily obtainable as paradata and is shown to be a consistent determinant of data quality. However, more detailed research may yield important lessons for survey providers in better utilizing paradata and developing supervision tools to ensure that data quality goals are met.

References

Adida, C. L., Ferree, K. E., Posner, D. N., & Robinson, A. L. (2016). "Who's asking? Interviewer coethnicity effects in African survey data". Comparative Political Studies, 49(12): pp. 1630-1660.

Baird, S., Hamory, J., & Miguel, E. (2008). "Tracking, attrition and data quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)". Center for International and Development Economics Research.

Bardasi E, Beegle K, Dillon A., & Serneels P. (2011). "Do labor statistics depend on how and to whom the questions are asked? Results from a survey experiment in Tanzania". World Bank Economic review, 25(3): pp. 418-447.

Barge, S., & Gehlbach, H. (2012). "Using the theory of satisficing to evaluate the quality of survey data." Research in Higher Education, 53(2): pp. 182-200.

Barth, A., & Schmitz, A. (2021). "Interviewers' and Respondents' Joint Production of Response Quality in Open-ended Questions. A Multilevel Negative-binomial Regression Approach". Methods, data, analyses, 15(1): pp. 43-76.

Beegle, K., Carletto, C., & Himelein, K. (2012). “Reliability of recall in agricultural data”. Journal of Development Economics, 98(1): pp. 34-41.

Biemer, P. P. (2010). “Total survey error: Design, implementation, and evaluation”. Public Opinion Quarterly, 74(5): pp. 817-848.

Booth, A. (2019). “Measuring poverty and income distribution in Southeast Asia”. Asian-Pacific Economic Literature, 33(1): pp. 3-20.

Caeyers, B., Chalmers, N., & De Weerdt, J. (2012). “Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment”. Journal of Development Economics, 98(1): pp. 19-33.

Campanelli, P., & O'Muircheartaigh, C. (1999). “Interviewers, interviewer continuity, and panel survey nonresponse”. Quality and quantity, 33(1): pp. 59-76.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). “Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior”. Public Opinion Quarterly, 60(3): pp. 345-375.

Costa, J., Paul T., & McCrae, R. R. (1997). “Personality trait structure as a human universal”. American Psychologist, 52: pp. 587–596.

Couper, M. P., & Groves, R. M. (1992). “The role of the interviewer in survey participation”. Survey Methodology, 18 (2): pp. 263-277.

Couper, M. P., & Hansen, S. E. (2002). “Computer-assisted interviewing”. Handbook of interview research: pp. 557-575.

Couper, M. P., & Kreuter, F. (2013). “Using paradata to explore item level response times in surveys”. Journal of the Royal Statistical Society: Series A (Statistics in Society), 176(1): pp. 271-286.

Dang, H. A. (2012). “Vietnam: A Widening Poverty Gap for Ethnic Minorities”. In G. Hall & H. Patrinos (Eds.). Indigenous Peoples, Poverty and Development (pp. 304-343). New York, NY: Cambridge University Press.

Dang, H. A. H., & Carletto, C. (2018). “The Seemingly Underappreciated Role of Panel Data in Measuring Poverty and Economic Transformation”. World Economics, 19(3), 45-60.

De Leeuw, E. D., Hox, J. J., & Snijkers, G. (1995). “The effect of computer-assisted interviewing on data quality. A review”. Market Research Society. Journal., 37(4): pp. 1-19.

Feskens, R., Hox, J., Lensvelt-Mulders, G., & Schmeets, H. (2006). “Collecting data among ethnic minorities in an international perspective”. Field methods, 18(3): pp. 284-304.

Fisher, M., Reimer, J. J., & Carr, E.R. (2010). “Who should be interviewed in surveys of household income?”. World Development, 38(7): pp. 966-973.

Fowler Jr, F. J., & Mangione, T. W. (1990). “Standardized survey interviewing: Minimizing interviewer-related error” (Vol. 18). Sage.

Galesic M., & Bosnjak, M. (2009). “Effects of questionnaire length on participation and indicators of response quality in a web survey”. Public opinion quarterly, 73(2): pp. 349-360.

Gibson, J. (2016). “Poverty measurement: we know less than policy makers realize”. Asia & the Pacific Policy Studies, 3(3): pp. 430-442.

Glewwe, P., & Dang, H. A. H. (2008). “The impact of decentralized data entry on the quality of household survey data in developing countries: Evidence from a randomized experiment in Vietnam”. The World Bank Economic Review, 22(1): pp. 165-185.

Grosh, M. & Glewwe, P. (2000). “Designing household survey questionnaires for developing countries.”. Washington, DC: World Bank.

Groves, R. M. (1989). “Survey errors and survey costs”. New York, NY: Wiley-Interscience.
Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011) “Survey methodology (Vol. 561)”. Hoboken, NJ: John Wiley and Sons.

Hardeweg, B., Klasen, S., & Waibel, H. (2013). “Establishing a database for vulnerability assessment”. In Vulnerability to Poverty (pp. 50-79). Palgrave Macmillan, London.

Jäckle, A., Lynn, P., Sinibaldi, J. & Tipping, S. (2013). “The effect of interviewer experience, attitudes, personality and skills on respondent co-operation with face-to-face surveys”. Survey Research Methods 7(1): pp. 1–15.

Kahn, R L., & Cannell C. F. (1957). “The dynamics of interviewing”. New York, NY: John Wiley and Sons.

Knäuper, B., Belli, R. F., Hill, D. H., & Herzog, A. R. (1997). “Question difficulty and respondents' cognitive ability: The effect on data quality”. Journal of Official Statistics-Stockholm, 13: pp. 181-199.

Kreuter, F., Couper, M., & Lyberg, L. (2010). “The use of paradata to monitor and manage survey data collection”. In Proceedings of the joint statistical meetings, American Statistical Association (pp. 282-296). Alexandria, VA: American Statistical Association.

Krosnick, J. A., & Alwin, D. F. (1987). “An evaluation of a cognitive theory of response-order effects in survey measurement”. Public Opinion Quarterly, 51(2): pp. 201-219.

Krosnick, J. A. (1991). “Response strategies for coping with the cognitive demands of attitude measures in surveys”. Applied Cognitive Psychology, 5: pp. 213-236.

Krumpal, I. (2013). “Determinants of social desirability bias in sensitive surveys: a literature review”. Quality & Quantity, 47(4): pp. 2025-2047.

Long, J. S., & Freese, J. (20149. “Regression Models for Categorical Dependent Variables Using Stata, 3rd Edition”. College Station, TX: Stata Press

Lupu, N., & Michelitch, K. (2018). “Advances in Survey Methods for the Developing World”. Annual Review of Political Science, Vol. 21, pp. 195-214.

Lynn, P., & Clarke, P. (2002). “Separating refusal bias and non-contact bias: evidence from UK national surveys”. Journal of the Royal Statistical Society: Series D (The Statistician), 51(3): pp. 319-333.

Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). “Household Surveys in Crisis”. Journal of Economic Perspectives, 29(4), 199-226.

Meyer, B. D., Mittag, N., & Goerge, R. M. (2018). “Errors in survey reporting and imputation and their effects on estimates of food stamp program participation”. (No. w25143). National Bureau of Economic Research.

Olson, K., & Peytchev, A. (2007). “Effect of interviewer experience on interview pace and interviewer attitudes”. Public Opinion Quarterly, 71(2): pp. 273-286.

Olson, K. & Bilgen, I. (2011). “The role of interviewer experience on acquiescence”. Public Opinion Quarterly, 75(1): pp. 99-114.

Olson, K., Kirchner, A., & Smyth, J. (2016). “Do interviewers with high cooperation rates behave differently? Interviewer cooperation rates and interview behaviors”. Survey Practice, 9(2): pp. 1-11.

Pennell, B. E., Cibelli Hibben, K. L., Lyberg, L., Mohler, P. P., & Worku, G. (2017). “A total survey error perspective on surveys in multinational, multiregional, and multicultural contexts”. Total Survey Error in Practice: pp. 179-202.

Phung, T. D., Hardeweg, B., Praneetvatakul, S., & Waibel, H. (2015). “Non-sampling error and data quality: What can we learn from surveys to collect data for vulnerability measurements?”. World Development, 71: pp. 25-35.

Singer, E., Frankel, M., & Glassman, M. (1983). “The effect of interviewer characteristics and expectation on response”. Public Opinion Quarterly, 47 (1): pp. 68-83.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). “The Psychology of Survey Response”. New York: Cambridge University Press.

Townsend, R. M., Sakunthasathien, S., & Jordan, R. (2013). “Chronicles from the field: The Townsend Thai Project”. Cambridge. MA: MIT Press.

Weisberg, H.F. (2005). “The total survey error approach: A guide to the new science of survey research”. Chicago, IL: University of Chicago Press.

Wollburg, P., Tiberti, M., & Zezza, A. (2021). “Recall length and measurement error in agricultural surveys”. Food Policy, 100, 102003.

Yu, L. (2012). “Using negative binomial regression analysis to predict software faults: a study of apache ant”. Information Technology and Computer Science, 4(8): pp. 63–70.

Appendix

	Combined regression model			Country-level regression model					
				Thailand			Vietnam		
	Missing data	Refusals	Implausible values	Missing data	Refusals	Implausible values	Missing data	Refusals	Implausible values
AIC	24633	14734	27890	12707	11474	6090	8297	14125	13427
Countfit preferred model	NBRM	NBRM	NBRM	NBRM	NBRM	NBRM	NBRM	NBRM	NBRM
Cragg-Uhlert/ Nahelkerke r^2	0.16	0.14	0.28	0.14	0.16	0.15	0.25	0.23	0.26
lnalpha significance	6.10E+04*	7074*	1.50E+04*	4.20E+04*	1.30E+04*	3167*	2526*	7587*	5887*

Note: The countfit command by Long & Freese (2014) compares the fit of Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial models and in doing so provides a table of estimates and the results of tests/measures on goodness of fit. Thus, the output specifies which count model is preferred by comparing each individual count model with one another.
* Significant at 1%.

Figure A1. Results of comparisons of fit for combined and country-level negative binomial models

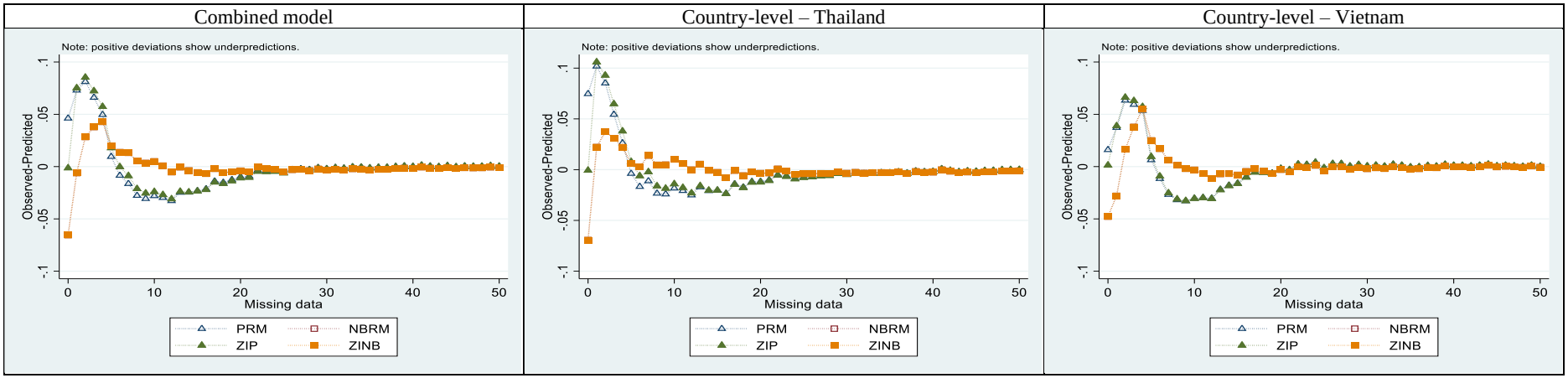

Figure A2. Graphical depiction of countfit results – Missing data

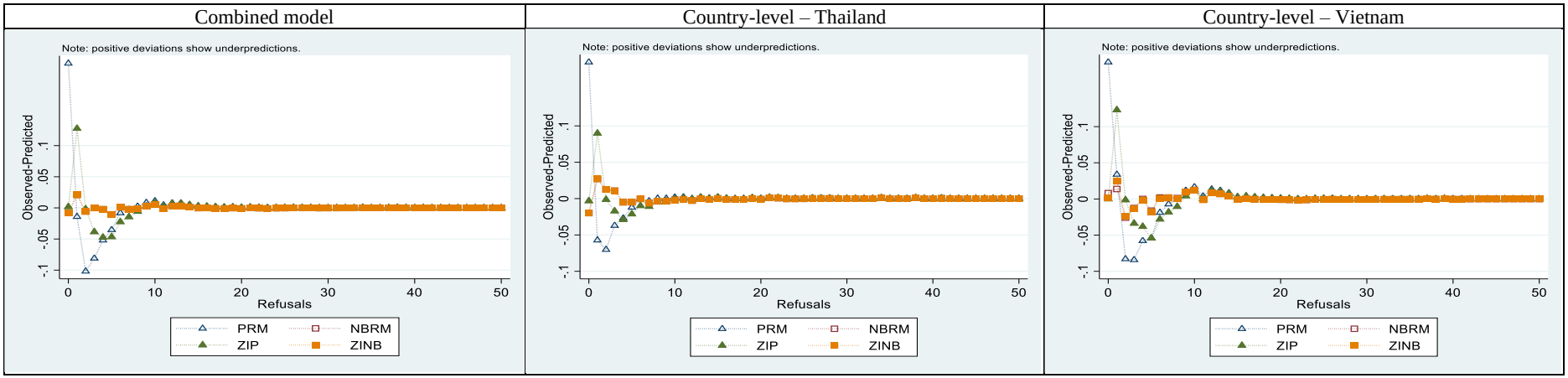

Figure A3. Graphical depiction of countfit results – Refusal

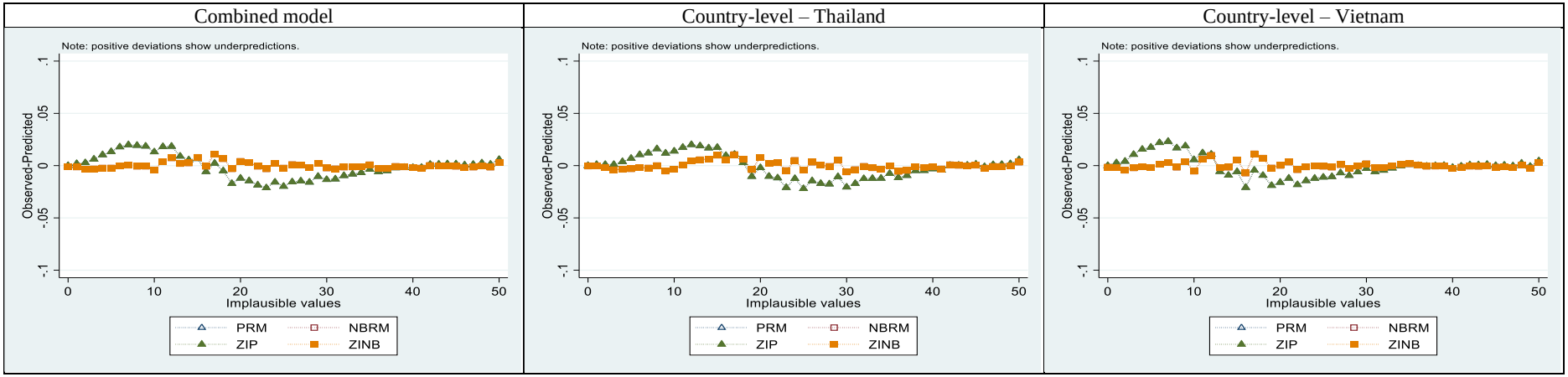

Figure A4. Graphical depiction of countfit results – Implausible value

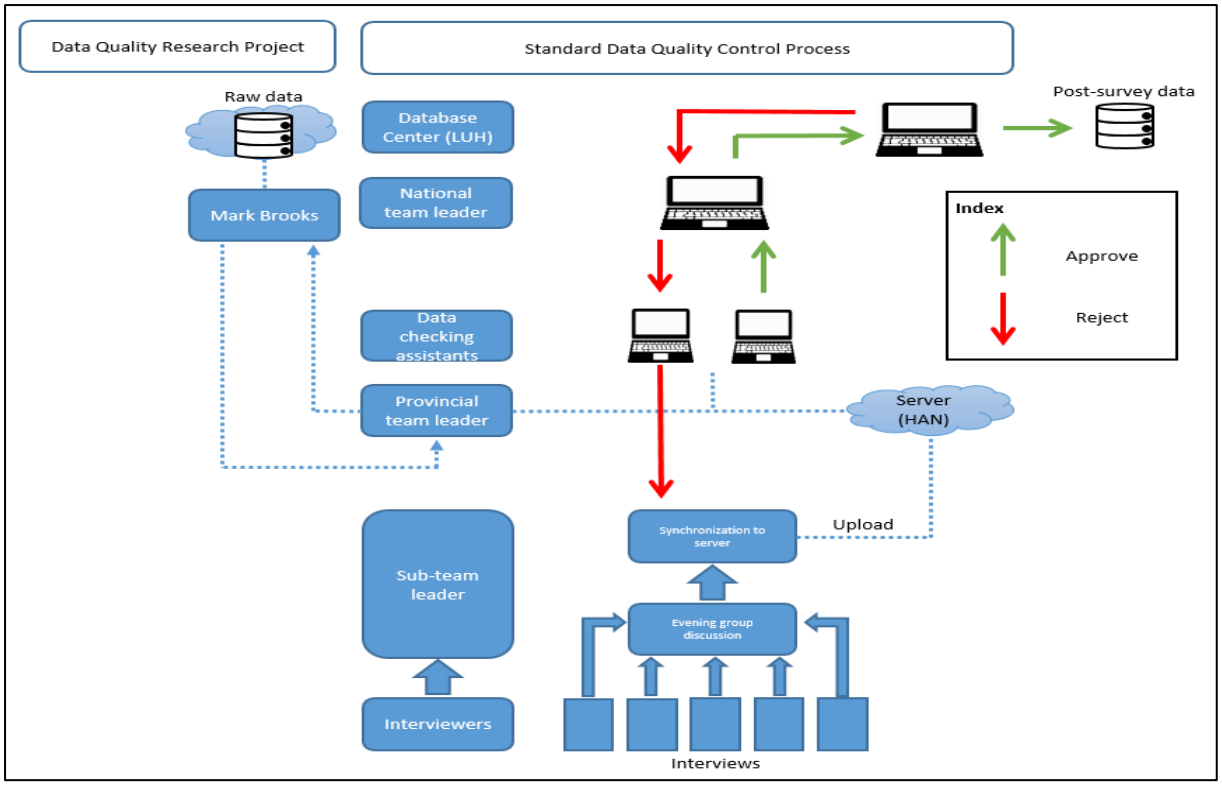

Figure A5. Survey data collection procedure – Example with one survey team

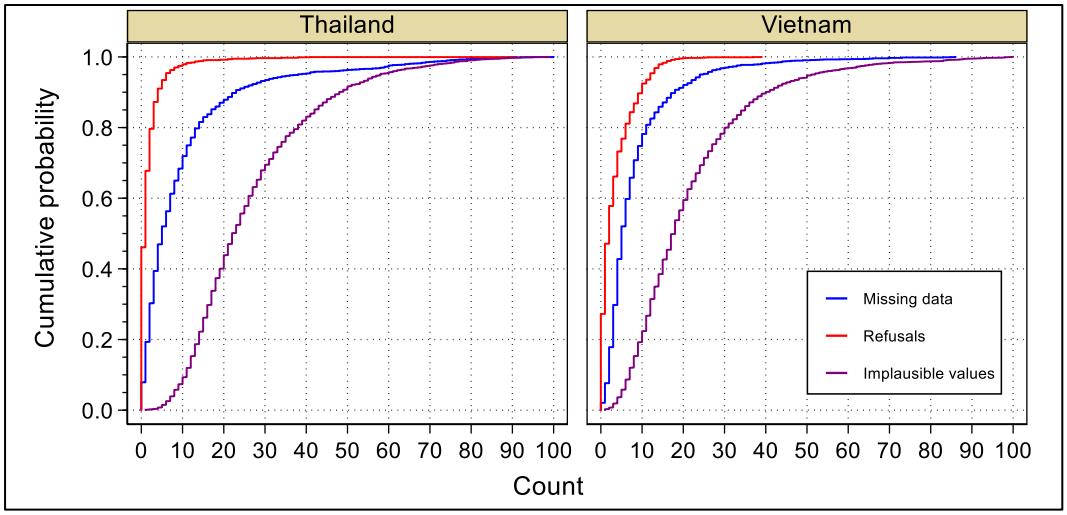

Figure A6. Cumulative distribution function of non-sampling errors, by country

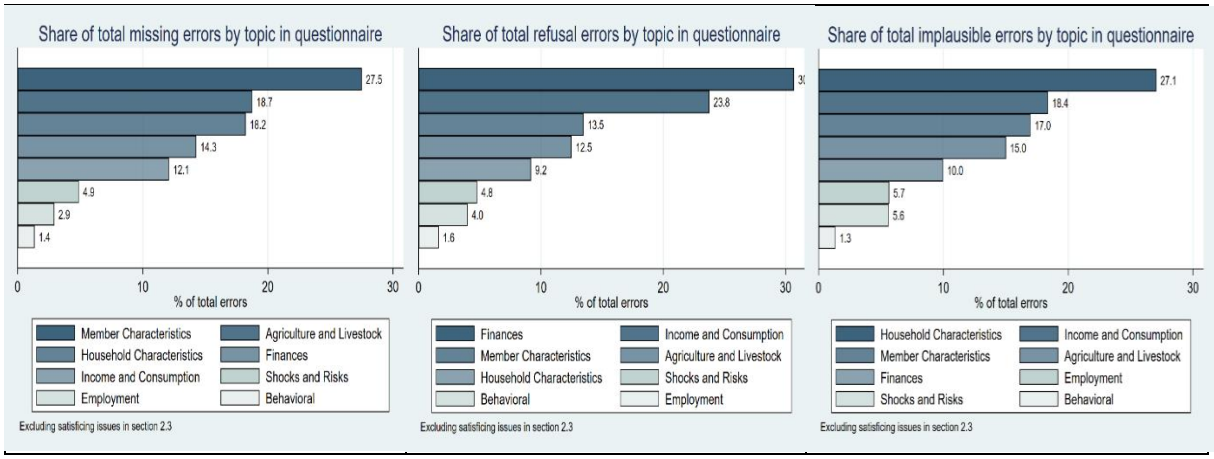

Figure A7. Overview of non-sampling errors by survey instrument module

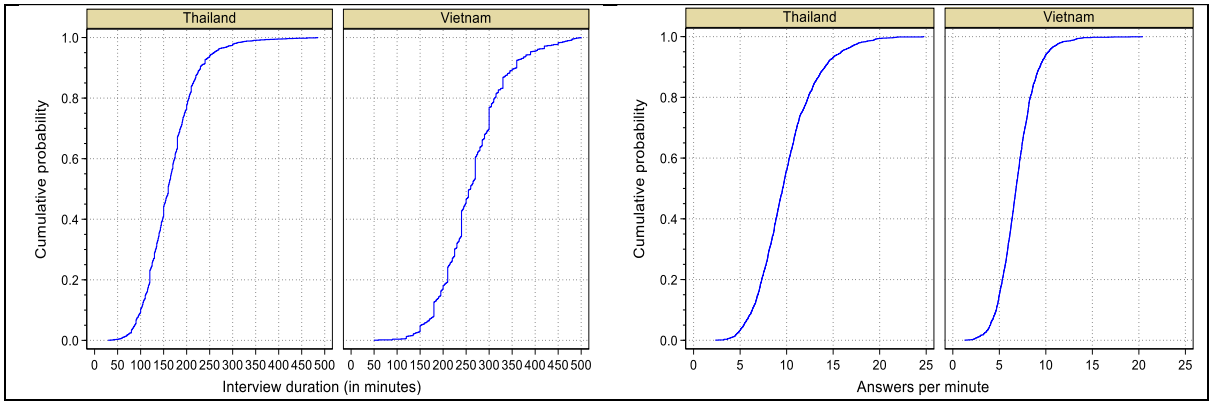

Figure A8. Cumulative probability function of interview duration and entry speed, by country

nbreg (N=1806): Percentage change in expected count

Observed SD: 43.9497

	b	z	P>|z|	%	%StdX	SDofX
1.Resp_Gen~r	-0.0161	-0.206	0.837	-1.6	-0.8	0.475
Resp_Age	-0.0047	-0.265	0.791	-0.5	-5.8	12.745
Resp_Age~e	0.0001	0.664	0.506	0.0	15.9	1468.752
1.Resp_Eth~t	0.3185	1.675	0.094	37.5	5.8	0.176
1.Resp_Hig~2	-0.0533	-0.543	0.587	-5.2	-1.9	0.363
Int_Gender						
Male	0.2635	2.740	0.006	30.2	12.6	0.450
Int_Age	0.0287	1.339	0.181	2.9	5.9	1.990
Int_Yeduc	-0.0581	-1.551	0.121	-5.6	-6.7	1.192
1.Int_Ethn~t	-0.2917	-1.640	0.101	-25.3	-5.6	0.197
Int_Local						
Born Locally	0.1102	1.392	0.164	11.7	5.6	0.498
Int_Openne~m	0.0578	0.916	0.360	6.0	3.9	0.666
Int_FoS						
1	-0.4354	-4.522	0.000	-35.3	-16.5	0.415
3	-0.2762	-3.620	0.000	-24.1	-12.6	0.489
1.Resp_Gen~r	0.1725	1.207	0.228	18.8	5.3	0.298
1.Int_Exp2~p	-0.0047	-0.110	0.913	-0.5	-0.4	0.798
2.Int_Exp2~p	-0.1644	-3.267	0.001	-15.2	-10.8	0.695
Inter_Dura~n	0.0010	1.106	0.269	0.1	6.1	56.647
answ_per_min	0.0160	1.132	0.258	1.6	5.8	3.521
1.FirstInt~w	-0.1126	-1.832	0.067	-10.7	-5.5	0.499
1.Inter_Ma~3	0.1276	1.605	0.109	13.6	5.5	0.418
1.Inter_Ot~s	-0.0333	-0.432	0.666	-3.3	-1.4	0.408
Survey_Week	-0.2414	-8.704	0.000	-21.4	-24.0	1.139
prov						
Ubon RatchaThai	0.1110	1.216	0.224	11.7	5.7	0.499
Nakhon Panom	0.4053	3.415	0.001	50.0	16.2	0.371
HHSize	0.0473	2.427	0.015	4.8	9.5	1.915
HH_AgriLan~e	-0.0002	-0.142	0.887	-0.0	-0.5	26.473
YIncCap1000	0.0009	0.321	0.748	0.1	0.5	6.049
constant	-4.4855	-5.303	0.000	.	.	.
alpha						
lnalpha	0.4094	.	.	.	.	.
alpha	1.5059	.	.	.	.	.

LR test of alpha=0: 4.2e+04 Prob>=LRX2 = 0.000
b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
% = percent change in expected count for unit increase in X
%StdX = percent change in expected count for SD increase in X
SDofX = standard deviation of X

nbreg (N=1827): Percentage change in expected count

Observed SD: 24.2723

	b	z	P>|z|	%	%StdX
1.Resp_Gen~r	0.0144	0.256	0.798	1.5	0.7
Resp_Age	-0.0078	-0.816	0.414	-0.8	-10.3
Resp_Age~e	0.0001	1.041	0.298	0.0	15.1
1.Resp_Hig~2	0.1347	2.665	0.008	14.4	6.7
Int_Age	0.0255	2.659	0.008	2.6	6.1
Int_Gender					
Male	-0.0188	-0.292	0.770	-1.9	-0.9
Int_Yeduc	-0.0523	-2.953	0.003	-5.1	-6.4
Int_Local					
Born Locally	0.2936	4.510	0.000	34.1	12.7
Int_Openne~m	0.0704	1.554	0.120	7.3	4.2
Int_FoS					
1	0.0021	0.037	0.971	0.2	0.1
3	-0.1215	-1.030	0.303	-11.4	-2.6
1.Resp_Eth~t	0.4589	1.528	0.126	58.2	20.7
1.Int_Ethn~t	-0.2604	-0.961	0.336	-22.9	-4.0
1.Resp_Eth~t	-0.6288	-2.084	0.037	-46.7	-23.3
1.Resp_Gen~r	0.1481	1.694	0.090	16.0	5.8
1.Int_Exp2~p	0.0137	1.301	0.193	1.4	3.5
2.Int_Exp2~p	-0.1600	-3.452	0.001	-14.8	-7.9
Inter_Dura~n	0.0000	0.079	0.937	0.0	0.2
answ_per_min	-0.0021	-0.133	0.894	-0.2	-0.4
1.FirstInt~w	-0.1308	-3.026	0.002	-12.3	-6.2
1.Inter_Ma~3	-0.1271	-2.154	0.031	-11.9	-4.8
1.Inter_Ot~s	-0.0734	-1.113	0.266	-7.1	-2.4
Survey_Week	-0.1412	-6.724	0.000	-13.2	-19.2
prov					
Thua Thien Hue	0.3022	3.713	0.000	35.3	14.9
Dak Lak	-0.3108	-4.332	0.000	-26.7	-13.8
HHSize	0.0070	0.467	0.640	0.7	1.3
HH_AgriLan~e	0.0007	0.849	0.396	0.1	2.1
YIncCap1000	0.0194	3.864	0.000	2.0	8.7
constant	-4.2563	-7.228	0.000	.	.
alpha					
lnalpha	-0.4402	.	.	.	.
alpha	0.6439	.	.	.	.

LR test of alpha=0: 1.3e+04 Prob>=LRX2 = 0.000
b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
% = percent change in expected count for unit increase in X
%StdX = percent change in expected count for SD increase in X
SDofX = standard deviation of X

Figure A9. Results of listcoef for missing data model – Thailand (left) and Vietnam (right)

nbreg (N=1806): Percentage change in expected count

Observed SD: 5.9469

	b	z	P>|z|	%	%StdX	SDofX
1.Resp_Gen~r	-0.0628	-0.599	0.549	-6.1	-2.9	0.475
Resp_Age	-0.0281	-1.389	0.165	-2.8	-30.1	12.745
Resp_Age~e	0.0003	1.671	0.095	0.0	53.2	1468.752
1.Resp_Hig~2	0.0630	0.514	0.607	6.5	2.3	0.363
1.Resp_Eth~t	0.2666	1.028	0.304	30.6	4.8	0.176
HeadHouse						
Yes	0.0966	1.016	0.310	10.1	4.9	0.495
Resp_Inter~s	-0.0233	-1.041	0.298	-2.3	-4.1	1.816
Resp_Extra	-0.0803	-2.272	0.023	-7.7	-8.1	1.054
Resp_Neuro	0.0332	0.952	0.341	3.4	3.8	1.124
Int_Gender						
Male	-0.0145	-0.132	0.895	-1.4	-0.7	0.450
Int_Age	0.0984	3.992	0.000	10.3	21.6	1.990
Int_Yeduc	-0.0562	-1.274	0.203	-5.5	-6.5	1.192
1.Int_Ethn~t	-0.6934	-3.199	0.001	-50.0	-12.8	0.197
Int_Local						
Born Locally	0.1935	1.997	0.046	21.4	10.1	0.498
Int_Traini~f	-0.0799	-1.161	0.245	-7.7	-5.0	0.644
Int_Agree_m	-0.2508	-2.990	0.003	-22.2	-14.9	0.644
Int_FoS						
1	0.4187	3.666	0.000	52.0	19.0	0.415
3	0.2006	2.008	0.045	22.2	10.3	0.489
1.Resp_Gen~r	-0.1001	-0.572	0.568	-9.5	-2.9	0.298
1.Int_Exp2~p	0.3570	7.183	0.000	42.9	33.0	0.798
2.Int_Exp2~p	0.1061	1.680	0.093	11.2	7.7	0.695
Inter_Dura~n	-0.0001	-0.112	0.911	-0.0	-0.7	56.647
answ_per_min	0.0296	1.562	0.118	3.0	11.0	3.521
1.FirstInt~w	0.1274	1.714	0.087	13.6	6.6	0.499
1.Inter_Ma~3	0.0521	0.531	0.596	5.3	2.2	0.418
1.Inter_Ot~s	-0.0644	-0.665	0.506	-6.2	-2.6	0.408
Survey_Week	0.2710	7.768	0.000	31.1	36.2	1.139
prov						
Ubon RatchaThai	0.1374	1.351	0.177	14.7	7.1	0.499
Nakhon Panom	-0.1851	-1.199	0.230	-16.9	-6.6	0.371
HHSize	-0.0273	-1.094	0.274	-2.7	-5.1	1.915
HH_AgriLan~e	-0.0011	-0.744	0.457	-0.1	-2.8	26.473
YIncCap1000	0.0146	2.317	0.020	1.5	9.2	6.049
constant	-6.3187	-5.591	0.000	.	.	.
alpha						
lnalpha	0.5048	.	.	.	.	.
alpha	1.6567	.	.	.	.	.

LR test of alpha=0: 3167.05 Prob>=LRX2 = 0.000
b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
% = percent change in expected count for unit increase in X
%StdX = percent change in expected count for SD increase in X
SDofX = standard deviation of X

nbreg (N=1827): Percentage change in expected count

Observed SD: 4.2281

	b	z	P>|z|	%	%StdX	SDofX
1.Resp_Gen~r	-0.3584	-4.109	0.000	-30.1	-16.3	0.496
Resp_Age	-0.0422	-3.317	0.001	-4.1	-44.3	13.858
Resp_Age~e	0.0005	3.994	0.000	0.0	100.7	1533.658
1.Resp_Hig~2	-0.2346	-3.561	0.000	-20.9	-10.6	0.479
HeadHouse						
Yes	0.0818	0.926	0.355	8.5	4.1	0.495
Resp_Inter~s	-0.0145	-0.682	0.495	-1.4	-2.5	1.746
Resp_Extra	-0.0338	-1.331	0.183	-3.3	-3.6	1.096
Resp_Neuro	0.0292	1.048	0.295	3.0	3.2	1.074
Int_Age	-0.0401	-2.839	0.005	-3.9	-8.9	2.332
Int_Gender						
Male	-0.5893	-6.576	0.000	-44.5	-25.1	0.490
Int_Yeduc	0.0356	1.494	0.135	3.6	4.6	1.264
Int_Local						
Born Locally	-0.2479	-2.998	0.003	-22.0	-9.6	0.407
Int_Traini~f	-0.0445	-0.832	0.405	-4.4	-2.6	0.587
Int_Agree_m	-0.2538	-4.779	0.000	-22.4	-13.9	0.591
Int_FoS						
1	0.2816	3.680	0.000	32.5	15.1	0.500
3	-0.0996	-0.616	0.538	-9.5	-2.2	0.220
1.Resp_Eth~t	0.1421	0.381	0.703	15.3	6.0	0.410
1.Int_Ethn~t	-0.5577	-1.667	0.096	-42.7	-8.4	0.157
1.Resp_Eth~t	-0.5977	-1.587	0.112	-45.0	-22.3	0.423
1.Resp_Gen~r	0.4185	3.570	0.000	52.0	17.3	0.381
1.Int_Exp2~p	-0.0501	-3.247	0.001	-4.9	-11.8	2.503
2.Int_Exp2~p	-0.1180	-1.877	0.060	-11.1	-5.9	0.516
Inter_Dura~n	-0.0017	-4.272	0.000	-0.2	-15.1	96.011
answ_per_min	-0.0847	-4.302	0.000	-8.1	-16.2	2.089
1.FirstInt~w	0.0424	0.741	0.459	4.3	2.1	0.492
1.Inter_Ma~3	-0.0824	-1.048	0.295	-7.9	-3.1	0.385
1.Inter_Ot~s	0.0294	0.334	0.738	3.0	1.0	0.327
Survey_Week	-0.0563	-2.022	0.043	-5.5	-8.1	1.508
prov						
Thua Thien Hue	0.7291	6.809	0.000	107.3	39.9	0.461
Dak Lak	0.7946	8.482	0.000	121.4	46.1	0.477
HHSize	-0.0104	-0.532	0.594	-1.0	-1.8	1.791
HH_AgriLan~e	0.0007	0.824	0.410	0.1	2.1	29.971
YIncCap1000	-0.0127	-1.726	0.084	-1.3	-5.3	4.318
constant	-0.7426	-0.795	0.426	.	.	.
alpha						
lnalpha	-0.0026	.	.	.	.	.
alpha	0.9974	.	.	.	.	.

LR test of alpha=0: 2525.62 Prob>=LRX2 = 0.000
b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
% = percent change in expected count for unit increase in X
%StdX = percent change in expected count for SD increase in X
SDofX = standard deviation of X

Figure A10. Results of listcoef for refusals model – Thailand (left) and Vietnam (right)

nbreg (N=1806): Percentage change in expected count

Observed SD: 18.9168

	b	z	P>|z|	%	%StdX	SDofX
1.Resp_Gen~r	0.0109	0.337	0.736	1.1	0.5	0.475
Resp_Age	-0.0069	-1.048	0.295	-0.7	-8.4	12.745
Resp_Age~e	0.0001	1.020	0.308	0.0	8.7	1468.752
1.Resp_Hig~2	-0.0059	-0.159	0.873	-0.6	-0.2	0.363
1.Resp_Eth~t	0.0489	0.662	0.508	5.0	0.9	0.176
HeadHouse						
Yes	0.0611	2.058	0.040	6.3	3.1	0.495
Resp_Inter~s	-0.0200	-2.820	0.005	-2.0	-3.6	1.816
Resp_Openn~s	-0.0002	-0.020	0.984	-0.0	-0.0	1.270
Int_Gender						
Male	-0.0667	-1.757	0.079	-6.5	-3.0	0.450
Int_Age	0.0107	1.344	0.179	1.1	2.2	1.990
Int_Yeduc	-0.0462	-3.186	0.001	-4.5	-5.4	1.192
1.Int_Ethn~t	-0.2318	-3.375	0.001	-20.7	-4.5	0.197
Int_Local						
Born Locally	0.2021	6.683	0.000	22.4	10.6	0.498
Int_Traini~f	-0.0700	-3.328	0.001	-6.8	-4.4	0.644
Int_Openne~m	0.0870	3.471	0.001	9.1	6.0	0.666
Int_Agree_m	-0.0785	-3.114	0.002	-7.6	-4.9	0.644
Int_Extra_m	-0.1161	-3.926	0.000	-11.0	-5.5	0.484
Int_FoS						
1	-0.2547	-6.951	0.000	-22.5	-10.0	0.415
3	-0.1738	-5.696	0.000	-16.0	-8.1	0.489
1.Resp_Gen~r	-0.0462	-0.840	0.401	-4.5	-1.4	0.298
1.Int_Exp2~p	-0.0058	-0.345	0.730	-0.6	-0.5	0.798
2.Int_Exp2~p	0.0119	0.609	0.542	1.2	0.8	0.695
Inter_Dura~n	0.0008	2.020	0.043	0.1	4.4	56.647
answ_per_min	0.0122	2.045	0.041	1.2	4.4	3.521
1.FirstInt~w	0.0218	0.923	0.356	2.2	1.1	0.499
1.Inter_Ma~3	0.0220	0.720	0.472	2.2	0.9	0.418
1.Inter_Ot~s	0.0195	0.655	0.512	2.0	0.8	0.408
Survey_Week	-0.1585	-14.815	0.000	-14.7	-16.5	1.139
prov						
Ubon RatchaThai	-0.0083	-0.239	0.811	-0.8	-0.4	0.499
Nakhon Panom	-0.0626	-1.257	0.209	-6.1	-2.3	0.371
HHSize	-0.0091	-1.157	0.247	-0.9	-1.7	1.915
HH_AgriLan~e	-0.0001	-0.112	0.911	-0.0	-0.1	26.473
YIncCap1000	0.0068	3.631	0.000	0.7	4.2	6.049
constant	-1.7196	-4.335	0.000	.	.	.
alpha						
lnalpha	-1.5915	.	.	.	.	.
alpha	0.2036	.	.	.	.	.

LR test of alpha=0: 7587.45 Prob>=LRX2 = 0.000
b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
% = percent change in expected count for unit increase in X
%StdX = percent change in expected count for SD increase in X
SDofX = standard deviation of X

nbreg (N=1827): Percentage change in expected count

Observed SD: 17.2798

	b	z	P>|z|	%	%StdX	SDofX
1.Resp_Gen~r	-0.0561	-1.446	0.148	-5.5	-2.7	0.496
Resp_Age	0.0082	1.452	0.147	0.8	12.0	13.858
Resp_Age~e	-0.0001	-1.022	0.307	-0.0	-7.6	1533.658
1.Resp_Eth~t	0.6316	3.420	0.001	88.1	29.5	0.410
Resp_Highe~2	0.0505	1.732	0.083	5.2	2.4	0.479
HeadHouse						
Yes	0.0817	2.092	0.036	8.5	4.1	0.495
Resp_Inter~s	-0.0202	-2.159	0.031	-2.0	-3.5	1.746
Resp_Openn~s	0.0134	1.408	0.159	1.4	1.9	1.375
Int_Age	-0.0467	-8.177	0.000	-4.6	-10.3	2.332
Int_Gender						
Male	0.1258	3.348	0.001	13.4	6.4	0.490
1.Int_Ethn~t	0.3252	1.913	0.056	38.4	5.2	0.157
Int_Yeduc	-0.0094	-0.916	0.360	-0.9	-1.2	1.264
Int_Local						
Born Locally	0.0185	0.501	0.616	1.9	0.8	0.407
Int_Traini~f	-0.1882	-7.866	0.000	-17.2	-10.5	0.587
Int_Openne~m	0.0397	1.424	0.155	4.1	2.3	0.578
Int_Agree_m	0.0608	2.394	0.017	6.3	3.7	0.591
Int_Extra_m	-0.0298	-0.757	0.449	-2.9	-1.1	0.374
Int_FoS						
1	0.0704	2.109	0.035	7.3	3.6	0.500
3	0.2045	2.788	0.005	22.7	4.6	0.220
1.Resp_Eth~t	-0.7135	-3.840	0.000	-51.0	-26.0	0.423
1.Resp_Gen~r	-0.0617	-1.240	0.215	-6.0	-2.3	0.381
1.Int_Exp2~p	0.0210	2.985	0.003	2.1	5.4	2.503
2.Int_Exp2~p	-0.1304	-4.806	0.000	-12.2	-6.5	0.516
Inter_Dura~n	0.0010	4.992	0.000	0.1	9.8	96.011
answ_per_min	0.0339	3.548	0.000	3.4	7.3	2.089
1.FirstInt~w	0.0122	0.493	0.622	1.2	0.6	0.492
1.Inter_Ma~3	-0.1039	-3.112	0.002	-9.9	-3.9	0.385
1.Inter_Ot~s	0.0973	2.539	0.011	10.2	3.2	0.327
Survey_Week	-0.0962	-7.926	0.000	-9.2	-13.5	1.508
prov						
Thua Thien Hue	-0.0251	-0.532	0.594	-2.5	-1.2	0.461
Dak Lak	-0.2514	-5.960	0.000	-22.2	-11.3	0.477
HHSize	-0.0208	-2.309	0.021	-2.1	-3.7	1.791
HH_AgriLan~e	-0.0002	-0.509	0.611	-0.0	-0.6	29.971
YIncCap1000	0.0047	1.829	0.067	0.5	2.0	4.318
constant	-2.7734	-5.950	0.000	.	.	.
alpha						
lnalpha	-1.5718	.	.	.	.	.
alpha	0.2077	.	.	.	.	.

LR test of alpha=0: 5887.03 Prob>=LRX2 = 0.000
b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
% = percent change in expected count for unit increase in X
%StdX = percent change in expected count for SD increase in X
SDofX = standard deviation of X

Figure A11. Results of listcoef for implausible values model – Thailand (left) and Vietnam (right)