

# Comprehensive data quality studies as a component of poverty assessments

Mark Brooks (Leibniz University Hannover)

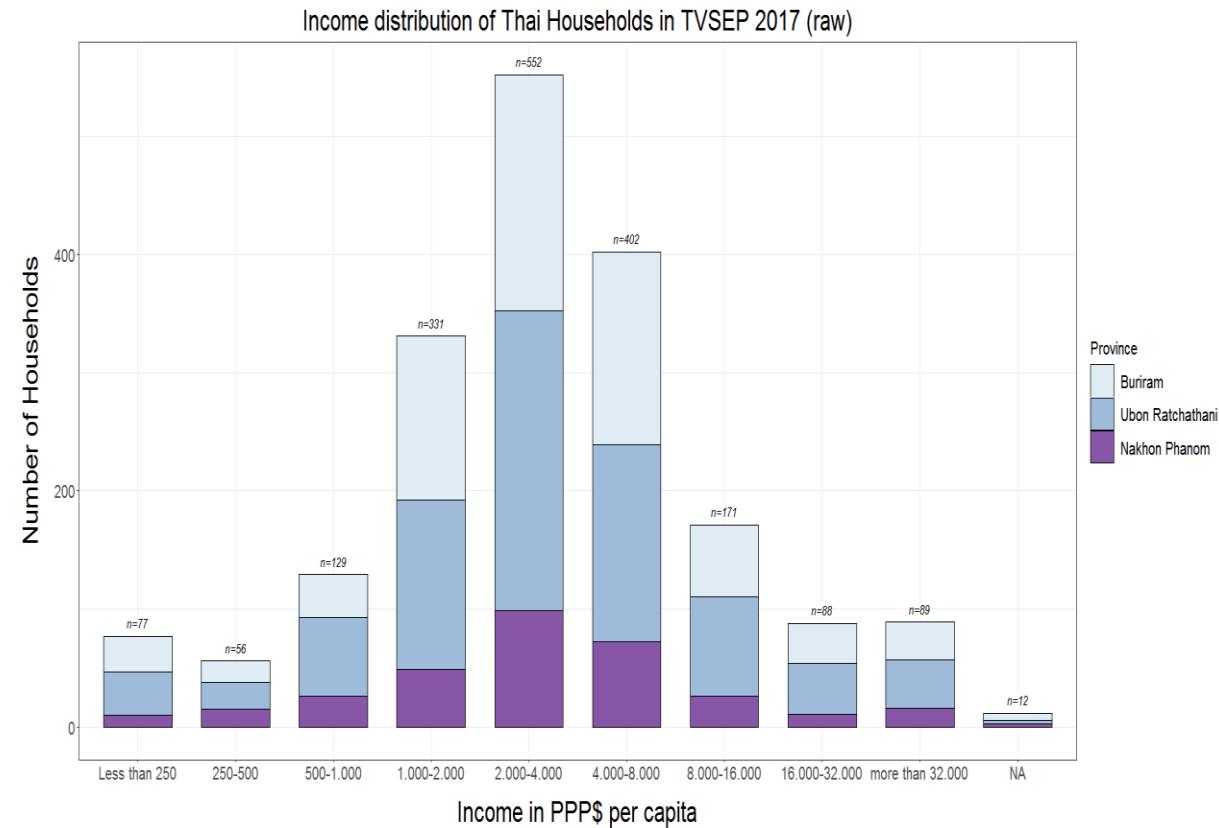
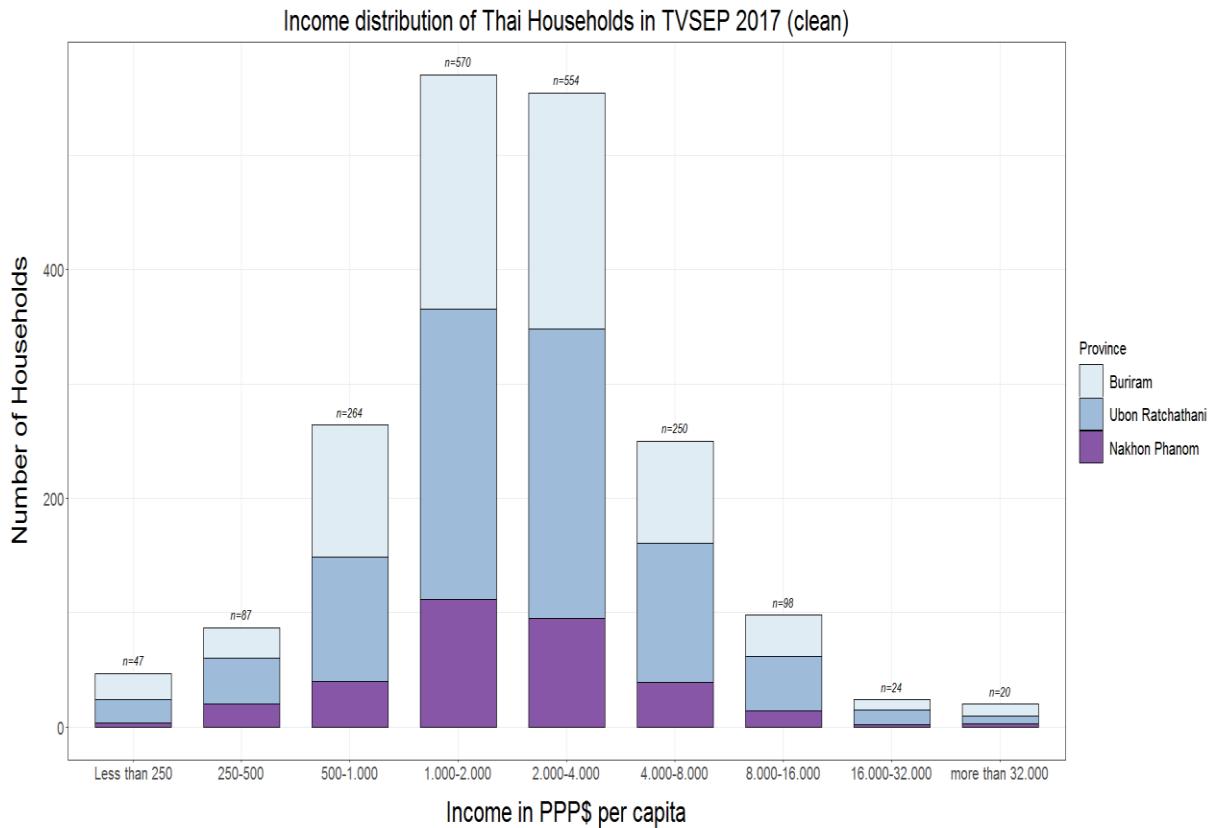
With Rattiya S. Lippe, Hermann Waibel

Special IARIW-World Bank Conference

Washington, DC – November 7-8, 2019

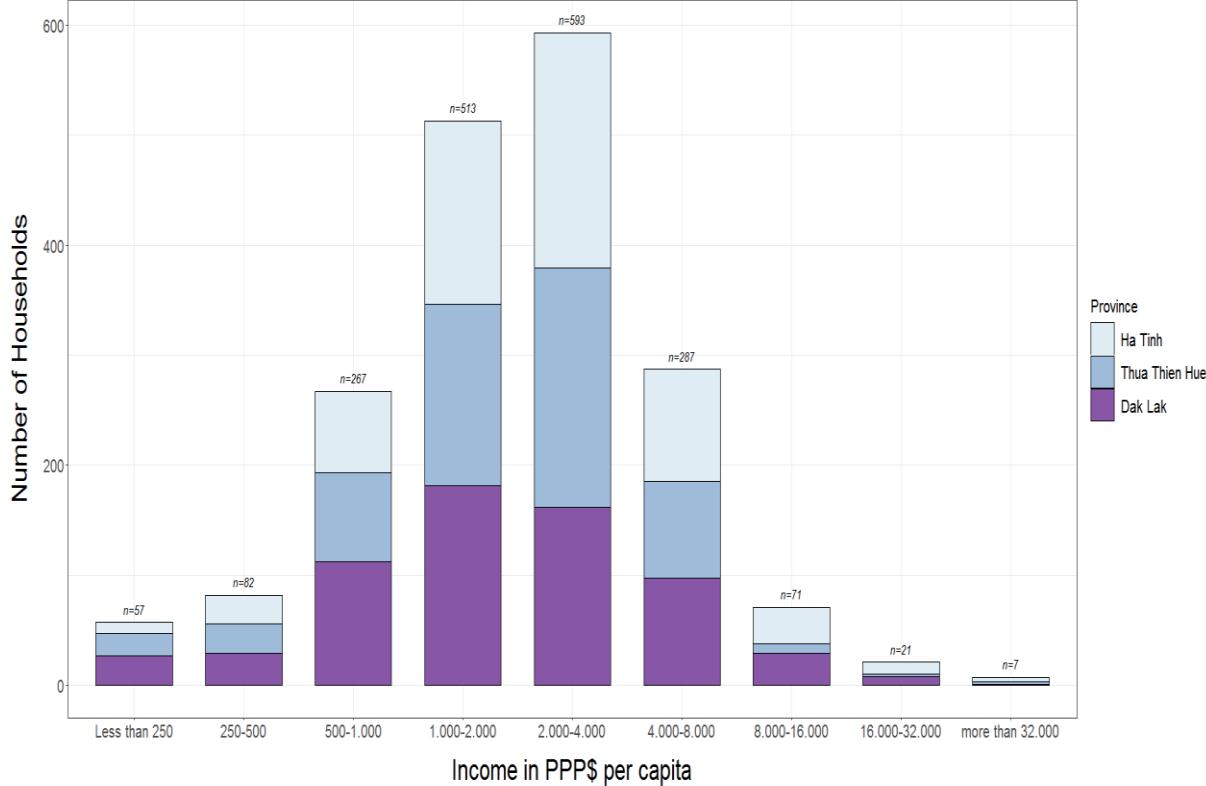
# Motivation (I)

- Household asset, income, and consumption data are the foundation for research on poverty
- Shortcomings of definitions of and data on household wealth may lead to misspecification of household poverty and result in incorrect policy conclusions

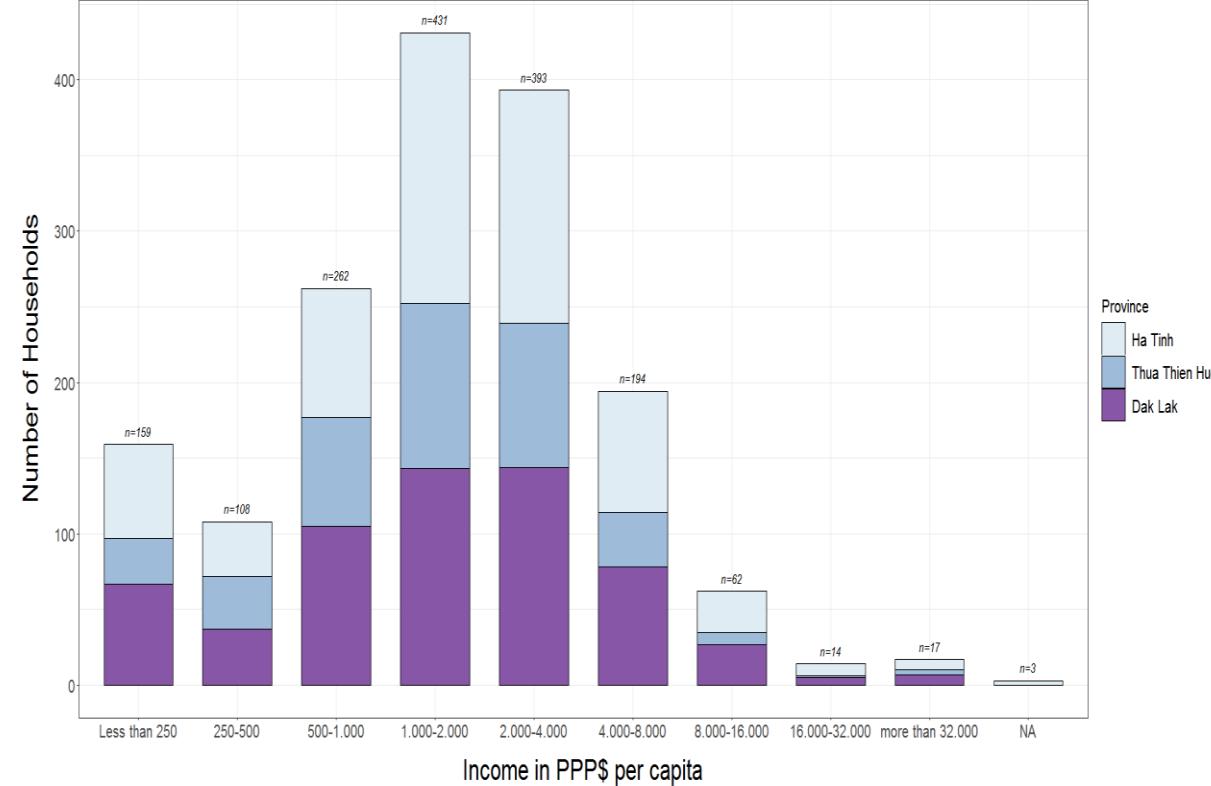


# Motivation (II)

Income distribution of Vietnamese Households in TVSEP 2017 (clean)



Income distribution of Vietnamese Households in TVSEP 2017 (raw)



→ Income on the low end of the distribution appears to frequently be underestimated in absence of supervision

# Constraints of Income Data & Objective

## Constraints of income data:

- Sensitive in nature, prone to recall bias, behavioral tendencies of respondents  
→ Susceptible to higher shares of non-sampling errors (e.g. Frick & Grabka, 2014; Meyer et al., 2018; Watson & Li, 2016)
- Few surveys suitable for use in calculating reliable poverty estimates (Booth, 2019; Gibson, 2016; Jolliffe et al., 2015)
- Advances in technology do not automatically solve the issue of low-quality data
- Quantity of high-quality survey data remains a constraint in developing countries (Booth, 2019; Dang & Carletto, 2018)

## Objective:

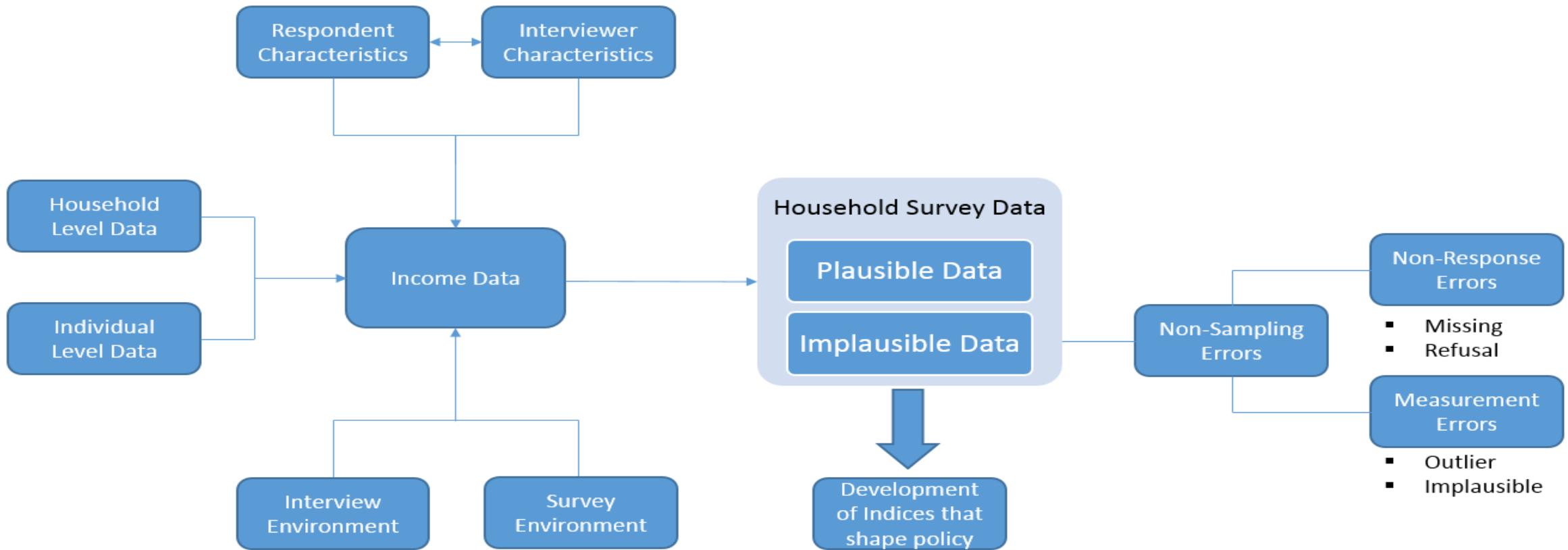
- Analyze the determinants of item non-response and measurement errors in income relevant data items

# Gaps in Research

- Most studies rely on cross-section data or experiments and are limited in scope regarding the types of non-sampling errors that can be studied
- Quantifiable interviewer and respondent characteristics have received much attention, whereas qualitative information (e.g. character traits) are seldom considered
- Studies seldom apply a comprehensive approach that allows for the identification of the relative importance of:
  - a) Interview environment;
  - b) Survey environment;
  - c) Interviewer and respondent characteristics;
  - d) Time.

## Conceptual Framework

Figure 1.  
Conceptual framework



*Source: Own representation*

# Non-Sampling Errors

- Consist of coverage errors, non-response errors, and measurement errors (Groves, 1989; Lessler & Kalsbeek, 1992; Weisberg, 2005)
- Non-Response Errors:
  - a) Unit non-response (excluded)
  - b) Item non-response
    - Missing values: When an item erroneously remains unanswered
    - Refusal value: When the interviewer ascertains that the respondent is unwilling to answer (TVSEP: signaled by choice of “no answer” option)
- Measurement errors:
  - a) Outlier values: Values outside of plausible ranges determined in the context of the survey
  - b) Implausible values: Answers that do not adhere to codelists or that in cross-reference with other related sections are found to be inconsistent

# Data

- Thailand Vietnam Socio Economic Panel (TVSEP) – 2017 Wave
- Target population: Rural households with livelihoods based in agriculture
- Sample size: 3,812 households
- Survey instrument: 10 sections, > 900 variables  
→ 84 variables examined related to components of income aggregate
- Use of “raw” data prior to data processing and with minimal supervision
- With addition of data quality modules sample is reduced to 3,447 households



Map source: Hardeweg (2011) based on ESRI World Map

# Model

1) Probit model → Capture the determinants of an interview containing a non-sampling error

$$Y_{in} = \alpha_0 + \beta_k * X_{ki} + \delta_m * Z_{mi} + \rho_n * F_{ni} + \eta_o * I_{oi} + \vartheta_p * S_{pi} + \lambda_{ji} * H_{ji} + \varepsilon_i \quad (1)$$

$Y_{in}$  - Dummy on prevalence of an error of type n (missing, refusal, outlier, implausible, total);

$X_{ki}$  - Interviewer characteristics;  $Z_{mi}$  - Respondent characteristics;  $F_{ni}$  - Congruent characteristics;

$I_{oi}$  - Interview environment;  $S_{pi}$  - Survey environment;  $H_{ji}$  - Household characteristics

2) OLS model → Capture the determinants of interviews containing higher shares of non-sampling errors

$$Y_{in} = \alpha_0 + \beta_k * X_{ki} + \delta_m * Z_{mi} + \rho_n * F_{ni} + \eta_o * I_{oi} + \vartheta_p * S_{pi} + \lambda_{ji} * H_{ji} + \varepsilon_i \quad (2)$$

$Y_{in}$  - Share of an error of type n (missing, refusal, outlier, implausible, total)

- Does not consider interviews found to be free of non-sampling errors

# Summary Statistics (I)

- In Thailand 9.4% and in Vietnam 14.5% of interviews were free of non-sampling errors and excluded in the OLS model

*Table 1. Summary statistics of errors by country*

Dependent variables *	Thailand			Vietnam		
	Obs.	Mean	Std. dev.	Obs.	Mean	Std. dev.
Missing values (in %)	1,818 (336)	0.94 (5.09)	4.69 (9.91)	1,629 (182)	0.27 (2.44)	1.41 (3.55)
Refusal values (in %)	1,818 (587)	0.65 (2.01)	2.25 (3.60)	1,629 (134)	0.08 (0.95)	0.38 (0.96)
Outlier values (in %)	1,818 (539)	0.43 (1.45)	0.95 (1.25)	1,629 (484)	0.23 (0.76)	0.50 (0.66)
Implausible values (in %)	1,818 (1,489)	3.81 (4.65)	4.38 (4.42)	1,629 (1,345)	2.38 (2.89)	3.09 (3.18)
Total erroneous values (in %)	1,818 (1,648)	5.83 (6.43)	7.04 (7.13)	1,629 (1,392)	2.96 (3.47)	3.71 (3.79)

Source: Own calculation based on TVSEP survey 2017 Notes: \*Number in brackets refers to descriptives for the OLS model

# Summary Statistics (II)

Table 2. Summary statistics of independent variables by country

Independent variables	Obs.		Mean		Std. dev.	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
<b>1) Respondent Characteristics</b>						
Respondent age (1 = $\geq 60$ ; 0 = $< 60$ )	1,818	1,629	0.42	0.26	0.49	0.44
Respondent gender (1 = male, 0 = female)	1,818	1,629	0.34	0.44	0.47	0.50
Respondent education – secondary and higher (1 = yes, 0 = no)	1,818	1,629	0.15	0.67	0.36	0.47
Respondent is head of household (1 = yes, 0 = no)	1,818	1,629	0.57	0.58	0.49	0.49
Respondent agreeableness	1,817	1,629	5.75	5.88	0.97	0.90
Respondent has agricultural background (1 = yes, 0 = no)	1,818	1,629	0.70	0.82	0.46	0.38
<b>2) Interviewer Characteristics</b>						
Interviewer gender (1 = male, 0 = female)	1,818	1,629	0.28	0.40	0.45	0.49
Interviewer has completed degree (1 = yes, 0 = no)	1,818	-	0.85	-	0.36	-
Interviewer from suitable field of study (1 = economics/agriculture, 0 = other)	1,818	1,629	0.22	0.48	0.42	19.11
Interviewer exam score (in %)	1,818	1,629	25.78	37.45	15.06	0.24
Interviewer household survey experience (1 = yes, 0 = no)	1,818	1,629	0.61	0.06	0.48	0.32
Interviewer other survey experience (1 = yes, 0 = no)	1,818	1,629	0.26	0.89	0.44	0.22
Interviewer TVSEP survey experience (1 = yes, 0 = no)	1,818	1,629	0.13	0.05	0.33	0.91
Interviewer openness	1,818	1,629	4.59	5.49	1.19	0.67
Interviewer extraversion	1,818	1,629	3.96	4.14	0.79	0.43

Source: Own calculation based on TVSEP survey 2017

# Summary Statistics (III)

Table 2. Summary statistics of independent variables by country (cont.)

Independent variables	Obs.		Mean		Std. dev.	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
<b>3) Congruent Characteristics:</b>						
Congruent ethnicity (1 = yes, 0 = no)	-	1,629	-	0.76	-	0.50
<b>4) Interview and Survey Environment</b>						
Interview duration (minutes)	1,817	1,629	165.43	274.18	59.08	86.69
Morning interview (1 = yes, 0 = no)	1,818	1,629	0.53	0.59	0.49	0.49
Respondent participated in all waves (1 = yes, 0 = no)	1,818	1,629	0.10	0.04	0.29	0.20
Household size (persons)	1,818	1,629	4.57	4.44	1.91	1.79

Source: Own calculation based on TVSEP survey 2017

# Results OLS – Main Findings (I)

Table 4. OLS model results by country

Independent variables	(1) Missing value		(2) Refusal value		(3) Outlier value		(4) Implausible value		(5) Total erroneous value	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
<b>1) Respondent Characteristics</b>										
Respondent age (1 = $\geq 60$ ; 0 = $< 60$ )	0.573 (1.544)	-0.293 (0.885)	0.728 (0.593)	-0.193 (0.375)	0.192 (0.169)	-0.157 (0.135)	0.050 (0.338)	-0.652** (0.328)	0.681 (0.557)	-0.832** (0.421)
Respondent is head of household (1 = yes, 0 = no)	0.350 (1.553)	-0.394 (0.898)	0.688*** (0.234)	0.168 (0.271)	0.070 (0.135)	-0.062 (0.089)	0.492 (0.302)	-0.026 (0.235)	0.813* (0.431)	-0.114 (0.286)
Respondent age (= 1) * head of household (Interaction effect)	1.400 (2.311)	1.912 (1.236)	-0.678 (0.809)	-0.032 (0.405)	-0.102 (0.228)	0.306** (0.151)	0.131 (0.465)	1.010** (0.422)	0.006 (0.769)	1.311** (0.509)
Respondent education – secondary and higher (1 = yes, 0 = no)	1.989 (1.832)	-0.207 (0.858)	0.298 (0.282)	0.003 (0.175)	0.373** (0.149)	0.045 (0.069)	-0.086 (0.312)	0.201 (0.203)	0.506 (0.521)	0.254 (0.250)
Respondent agreeableness	-1.274** (0.565)	-0.322 (0.236)	0.117 (0.208)	0.091 (0.113)	-0.018 (0.059)	0.006 (0.033)	-0.081 (0.145)	-0.076 (0.089)	-0.172 (0.255)	-0.084 (0.104)
Respondent has agricultural background (1 = yes, 0 = no)	-4.305*** (1.389)	-2.521* (1.282)	-1.086** (0.453)	-0.895*** (0.327)	-0.762*** (0.176)	-0.470*** (0.139)	-1.661*** (0.313)	-1.266*** (0.388)	-2.511*** (0.475)	-1.395*** (0.420)

# Results OLS – Main Findings (II)

Table 4. OLS model results by country

Independent variables	(1) Missing value		(2) Refusal value		(3) Outlier value		(4) Implausible value		(5) Total erroneous value	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
<b>2) Interviewer Characteristics</b>										
Interviewer gender (1 = male, 0 = female)	2.511*	-0.274	-0.163	-0.119	-0.112	0.023	-1.072***	0.810***	-0.568	0.769***
	(1.291)	(0.552)	(0.315)	(0.141)	(0.172)	(0.065)	(0.290)	(0.198)	(0.466)	(0.225)
Interviewer from suitable field of study (1 = economics/agriculture, 0 = other)	-1.850*	-0.096	0.347	-0.146	-0.284**	-0.066	-0.671**	0.401*	-1.077**	0.457*
	(1.024)	(0.608)	(0.696)	(0.164)	(0.110)	(0.065)	(0.320)	(0.233)	(0.447)	(0.255)
Interviewer exam score (in %)	0.096***	0.003	-0.008	-0.001	-0.009**	0.001	-0.058***	0.009	-0.046***	0.010
	(0.035)	(0.021)	(0.020)	(0.004)	(0.004)	(0.002)	(0.009)	(0.006)	(0.014)	(0.007)
Interviewer household survey experience (1 = yes, 0 = no) (No experience is base)	-0.882	-2.061	0.510	0.186	-0.235**	0.045	-0.132	-0.753**	-0.208	-1.069**
	(1.186)	(1.308)	(0.486)	(0.191)	(0.118)	(0.099)	(0.315)	(0.349)	(0.465)	(0.527)
Interviewer TVSEP survey experience (1 = yes, 0 = no) (No experience is base)	-1.696	-4.347***	0.103	0.132	0.227	-0.150	0.108	-0.384	0.041	-1.050
	(1.476)	(1.537)	(0.345)	(0.432)	(0.208)	(0.133)	(0.329)	(0.693)	(0.470)	(0.774)
Interviewer extraversion	2.762***	0.329	0.074	-0.116	-0.178***	-0.036	-0.258*	-0.084	0.194	-0.080
	(0.982)	(0.387)	(0.239)	(0.127)	(0.065)	(0.045)	(0.141)	(0.112)	(0.267)	(0.135)
<b>3) Congruent Characteristics</b>										
Congruent ethnicity (1 = yes, 0 = no)	Omit.	0.038	Omit.	-0.076	Omit.	0.150**	Omit.	-0.227	Omit.	-0.344
		(0.804)		(0.192)		(0.063)		(0.194)		(0.256)

# Results OLS – Main Findings (III)

Table 4. OLS model results by country

Independent variables	(1) Missing value		(2) Refusal value		(3) Outlier value		(4) Implausible value		(5) Total erroneous value	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
<b>4) Interview and Survey Environment</b>										
Morning interview (1 = yes, 0 = no)	-0.807 (0.983)	-0.010 (0.603)	0.085 (0.307)	-0.235 (0.153)	-0.011 (0.104)	0.119* (0.061)	0.469** (0.220)	-0.024 (0.161)	0.253 (0.349)	-0.039 (0.196)
Respondent participated in all waves (1 = yes, 0 = no)	-3.310** (1.573)	-0.763 (1.198)	-0.415 (0.471)	-0.232 (0.308)	0.168 (0.228)	-0.140 (0.137)	-0.001 (0.396)	0.071 (0.364)	-0.839 (0.517)	-0.012 (0.436)
Survey week	0.024 (0.431)	0.136 (0.224)	-0.089 (0.196)	-0.003 (0.075)	-0.137*** (0.047)	-0.003 (0.035)	-0.589*** (0.101)	-0.351*** (0.116)	-0.871*** (0.165)	-0.458*** (0.120)
Constant	5.897 (10.763)	2.118 (6.317)	10.308*** (3.385)	5.841** (2.310)	6.470*** (1.116)	2.581*** (0.723)	11.323*** (1.919)	9.575*** (2.908)	15.301*** (3.354)	9.043*** (3.051)
Observations	335	182	585	134	539	484	1,487	1,345	1,646	1,392
Adj. R <sup>2</sup>	0.120	0.086	0.059	0.227	0.143	0.107	0.108	0.070	0.069	0.061

\* Significant at 10%; \*\* Significant at 5%; \*\*\* Significant at 1%.

Notes: Robust standard errors in parentheses

Source: Own calculation based on TVSEP survey 2017

# Conclusion & Outlook

- Quantitative and qualitative characteristics of interviewers and respondents are the main drivers of non-sampling errors
- Early waves of TVSEP prone to outlier and missing values (e.g. Phung et al., 2015) – at present implausible errors represent greatest threat to reliable income data
- Homogeneous characteristics of interviewers in data represents a bottleneck to our analysis

## Planned additions:

- More complex variables (e.g. travel distance using GIS; variable capturing complexity of hh income)
- Analysis using Heckmann / Hurdle model approach
- Analysis of TVSEP consumption data
- Use of validation data may improve the process of identifying non-sampling errors in household survey data (e.g. Epland & Kirkland, 2002; Mathiowetz et al., 2002; Meyer et al., 2019)  
→ Rapid increase of cooperation between surveys in similar contexts

Thank you for your interest!

All comments and suggestions are warmly welcome:  
[brooks@ifgb.uni-hannover.de](mailto:brooks@ifgb.uni-hannover.de)

# References

- Booth, A. (2019). Measuring poverty and income distribution in Southeast Asia. *Asian-Pacific Economic Literature*.
- 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.
- Epland, J., & Kirkeberg, M. I. (2002), "Comparing Norwegian Income Data in Administrative Registers With Income Data in the Survey of Living Conditions,". In: The International Conference on Improving Surveys (ICIS), Copenhagen, August 25–28, 2002.
- Frick, J. R.; Grabka, M. M. (2014). "Missing income data in the German SOEP: Incidence, imputation and its impact on the income distribution". *SOEP Survey Papers*, No. 225, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.
- Groves, R. M. (1989). "Survey errors and survey costs". New York, NY: Wiley-Interscience.
- Hardeweg, B., Klasen, S., & Waibel, H. (2013). "Establishing a database for vulnerability assessment". In *Vulnerability to Poverty* (pp. 50-79). Palgrave Macmillan, London.
- Jolliffe, D., Lanjouw, P., Chen, S., Kraay, A., Meyer, C., Negre, M., Prydz, E., Vakis, R., & Wethli, K. (2015). "A measured approach to ending poverty and boosting shared prosperity: concepts, data, and the twin goals". The World Bank, Washington.
- Lessler, J. T., & Kalsbeek, W. (1992). "Nonresponse errors in surveys". New York, NY: John Wiley and Sons.
- Mathiowetz, N. A., Brown, C., & Bound, J. (2002). "Measurement Error in Surveys of the Low-Income Population". In *Studies of Welfare Populations: Data collection and research issues*, eds. Michele Ver Ploeg, Robert A. Moffitt, and Constance F. Citro: pp. 157-194. Washington, DC: National Academy Press.
- 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.
- Meyer, B. D., & Mittag, N. (2019). "Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness, and holes in the safety net". *American Economic Journal: Applied Economics*, 11(2): pp. 176-204.
- 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.
- Watson, N., & Li, N. (2016). "Evaluating potential improvements to the income imputation methods for HILDA Survey". HILDA Discussion Paper No. 1/16: Melbourne Institute for Applied Economic and Social Research, University of Melbourne.
- Weisberg, H.F. (2005). "The total survey error approach: A guide to the new science of survey research". Chicago, IL: University of Chicago Press.

# Additional slides

# Background Add-On Project (I)

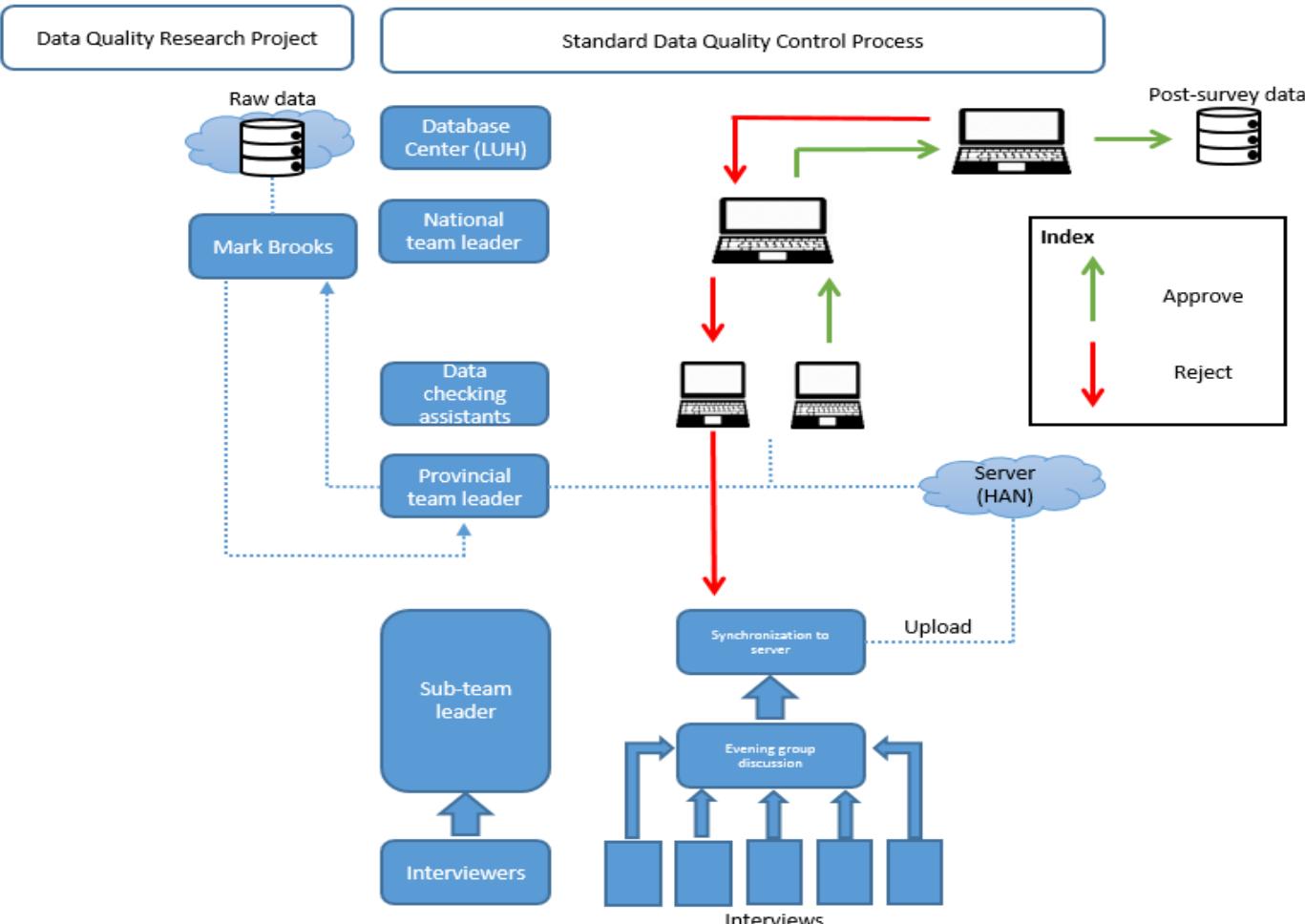
Figure 3.  
Data quality modules added to TVSEP 2017 survey



Source: Own representation

# Background Add-On Project (II)

Figure 2.  
Survey data structure TVSEP 2017 – Example with one team



Source: Own representation

# Sampling Process TVSEP

- Three-stage cluster sampling
- Target population: Purposive selection of 3 provinces that met required conditions
- First stage:
  - Probability proportional to size (PPS)
  - Provinces divided into strata with approximately proportional sample size
  - Systematic random sample with implicit stratification by population density
- Second stage:
  - Simple random PPS sample of two villages per sub-district
  - Sampled with a probability proportional to size of population from each sampled sub-district
- Third stage:
  - Equal probability of selection (EPS) used to draw 10 households from a list ordered by household size

# Results Probit (I)

Table 3. Probit model results by country

Independent variables	(1) Missing value		(2) Refusal value		(3) Outlier value		(4) Implausible value		(5) Total erroneous value	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
<b>Respondent age (1 = ≥ 60 ; 0 = &lt; 60)</b>	0.155 (1.21)	0.275 (1.42)	-0.004 (-0.03)	0.184 (0.087)	0.095 (0.82)	-0.012 (-0.07)	0.092 (0.67)	-0.192 (-1.13)	0.018 (0.11)	0.089 (0.48)
<b>Respondent is head of household (1 = yes, 0 = no)</b>	0.165 (1.67)	-0.009 (-0.07)	-0.042 (-0.46)	0.095 (0.69)	0.019 (0.21)	-0.005 (-0.05)	-0.072 (-0.69)	0.074 (0.65)	0.064 (0.50)	0.112 (0.94)
<b>Respondent age (= 1) * head of household (Interaction effect)</b>	-0.179 (-1.17)	-0.509* (-2.31)	-0.111 (-0.77)	-0.136 (-0.56)	-0.114 (-0.80)	0.080 (0.42)	-0.133 (-0.81)	0.391* (1.97)	-0.137 (-0.70)	0.091 (0.42)
<b>Respondent education – secondary and higher (1 = yes, 0 = no)</b>	0.057 (0.57)	0.094 (0.79)	0.075 (0.84)	-0.031 (-0.25)	0.236** (2.63)	0.174* (1.78)	-0.009 (-0.09)	0.273* (3.05)	0.101 (0.75)	0.268* (2.85)
<b>Respondent gender (1 = male, 0 = female)</b>	0.046 (0.57)	0.091 (0.77)	0.115 (1.48)	-0.031 (-0.25)	0.086 (1.11)	0.161 (1.80)	0.034 (0.39)	-0.129 (-1.26)	0.042 (0.42)	-0.142 (-1.33)
<b>Respondent agreeableness</b>	0.039 (1.02)	0.031 (0.61)	0.015 (0.45)	0.022 (0.46)	0.026 (0.76)	-0.034 (-0.90)	0.035 (0.91)	0.080 (1.77)	0.041 (0.95)	0.077 (1.62)
<b>Respondent has agricultural background (1 = yes, 0 = no)</b>	-0.036 (-0.45)	0.295* (2.28)	0.051 (0.67)	-0.024 (-0.19)	0.512*** (6.60)	0.372*** (3.79)	0.570*** (7.24)	0.796*** (8.30)	0.514*** (5.61)	0.791*** (8.02)
<b>Interviewer gender (1 = male, 0 = female)</b>	0.009 (0.10)	-0.018 (-0.19)	-0.038 (-0.46)	-0.085 (-0.83)	-0.374*** (-4.49)	0.015 (0.19)	-0.222* (-2.41)	0.153 (1.75)	-0.266* (-2.52)	0.123 (1.35)
<b>Interviewer has completed degree (1 = yes, 0 = no)</b>	-0.121 (-1.07)	Omit. (-4.49)	-0.455*** (-4.49)	Omit. (-1.77)	-0.189 (-1.77)	Omit. (-1.77)	0.210 (1.84)	Omit. (1.84)	-0.023 (-0.18)	Omit. (-0.18)
<b>Interviewer from suitable field of study (1 = economics/agriculture, 0 = other)</b>	-0.091 (-0.95)	0.038 (0.40)	-0.307*** (-3.40)	0.413*** (4.14)	-0.301*** (-3.34)	0.161* (2.07)	-0.088 (-0.92)	0.293*** (3.36)	-0.238* (-2.24)	0.274** (3.01)
<b>Interviewer exam score (in %)</b>	-0.002 (-0.52)	-0.000 (-0.15)	0.000 (0.00)	0.006* (2.15)	-0.007* (-2.48)	0.002 (1.11)	-0.009** (-2.85)	0.000 (0.04)	-0.006 (-1.67)	0.002 (0.67)
<b>Interviewer household survey experience (1 = yes, 0 = no) (No experience is base)</b>	-0.050 (-0.55)	-0.105 (-0.58)	0.084 (1.01)	0.796** (2.99)	-0.071 (-0.86)	0.250 (1.63)	0.017 (0.18)	0.439* (2.70)	0.013 (0.11)	0.524** (3.17)
<b>Interviewer TVSEP survey experience (1 = yes, 0 = no) (No experience is base)</b>	-0.123 (-1.00)	-0.647* (-1.99)	0.111 (1.03)	1.068** (3.33)	-0.120 (-1.13)	0.119 (0.52)	0.103 (0.85)	0.306 (1.25)	-0.134 (-0.96)	0.423 (1.67)
<b>Interviewer openness</b>	0.019 (0.54)	0.053 (0.99)	-0.064 (-1.95)	-0.190*** (-3.63)	0.000 (0.00)	0.074 (1.75)	0.029 (0.80)	-0.085 (-1.58)	0.015 (0.32)	-0.068 (-1.19)
<b>Interviewer extraversion</b>	-0.018 (-0.38)	-0.052 (-0.78)	-0.022 (-0.47)	0.066 (0.74)	0.041 (0.91)	-0.004 (-0.07)	-0.114* (-2.15)	-0.039 (-0.62)	-0.061 (-1.00)	-0.049 (-0.75)

# Results Probit (II)

Table 3. Probit model results by country (cont.)

Independent variables	(1) Missing value		(2) Refusal value		(3) Outlier value		(4) Implausible value		(5) Total erroneous value	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
Congruent ethnicity (1 = yes, 0 = no)	Omit. (-0.67)	-0.076 (-0.67)	Omit. (-1.06)	-0.130 (-1.06)	Omit. (3.33)	0.049 (0.53)	Omit. (-3.43)	-0.361*** (-3.43)	Omit. (-2.55)	-0.278* (-2.55)
Log of interview duration (minutes)	0.090 (0.84)	0.306* (2.06)	0.060 (0.61)	-0.012 (-0.07)	0.343*** (5.22)	0.615*** (5.22)	0.308** (2.74)	0.808*** (5.29)	0.424** (3.21)	0.821*** (5.15)
Morning interview (1 = yes, 0 = no)	-0.013 (-0.19)	-0.000 (-0.00)	-0.056 (-0.61)	0.016 (0.17)	0.097 (1.49)	0.018 (0.26)	0.131 (1.83)	0.066 (0.82)	0.036 (0.42)	0.100 (1.18)
Respondent participated in all waves (1 = yes, 0 = no)	-0.297* (-2.17)	0.215 (0.90)	-0.054 (-0.46)	-0.077 (-0.27)	-0.051 (-0.42)	0.035 (0.18)	-0.020 (-0.15)	-0.425* (-1.96)	-0.105 (-0.69)	-0.280 (-1.26)
Survey week	-0.163*** (-4.91)	-0.180*** (-4.00)	0.281*** (9.45)	-0.047 (-1.08)	-0.124*** (-4.24)	-0.076* (-2.20)	-0.054 (-1.66)	-0.065 (-1.68)	0.012 (0.30)	-0.091* (-2.24)
Buriram province (1 = yes, 0 = no) (Nakhon Phanom is base)	-0.182 (-1.58)	Omit. (-2.39)	-0.259* (-2.39)	Omit. (-0.60)	0.067 (0.60)	Omit. (-0.46)	-0.058 (-0.46)	Omit. (-0.14)	-0.020 (-0.14)	Omit. (-0.14)
Ubon province (1 = yes, 0 = no) (Nakhon Phanom is base)	-0.024 (-0.21)	Omit. (-0.04)	-0.004 (-0.04)	Omit. (-0.03)	-0.004 (-0.03)	Omit. (-0.38)	-0.046 (-0.38)	Omit. (0.46)	0.068 (0.46)	Omit. (0.46)
Ha Tinh province (1 = yes, 0 = no) (Dak Lak is base)	Omit. (0.13)	0.016 (0.13)	Omit. (-0.80)	-0.111 (-0.80)	Omit. (3.08)	0.310*** (3.08)	Omit. (2.94)	0.350* (2.94)	Omit. (2.87)	0.362* (2.87)
Hue province (1 = yes, 0 = no) (Dak Lak is base)	Omit. (0.62)	0.099 (0.62)	Omit. (-1.33)	-0.207 (-1.33)	Omit. (1.56)	0.188 (1.56)	Omit. (3.22)	0.448** (3.22)	Omit. (2.73)	0.386** (2.73)
Household size (persons)	0.025 (1.34)	0.018 (0.67)	0.036* (2.03)	0.102*** (3.72)	0.048** (2.74)	0.047* (2.16)	0.088*** (4.01)	0.103*** (3.79)	0.089*** (3.31)	0.119*** (4.09)
Constant	-1.106 (-1.73)	-2.818** (-2.93)	-0.883 (-1.53)	-1.966 (-1.79)	-2.480*** (-4.04)	-5.239*** (-6.62)	-0.980 (-1.43)	-4.919*** (-5.04)	-1.298 (-1.65)	-5.070*** (-4.97)
Observations	1,816	1,629	1,816	1,629	1,816	1,629	1,816	1,629	1,816	1,629
Pseudo R <sup>2</sup>	0.027	0.056	0.074	0.068	0.071	0.072	0.094	0.169	0.101	0.173

\* Significant at 10%; \*\* Significant at 5%; \*\*\* Significant at 1%.

Notes: Absolute value of z-statistics in parentheses

Source: Own calculation based on TVSEP survey 2017

# Results OLS (I)

Table 3. Probit model results by country  
Independent variables

	(1) Missing value		(2) Refusal value		(3) Outlier value		(4) Implausible value		(5) Total erroneous value	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
<b>Respondent age (1 = ≥ 60 ; 0 = &lt; 60)</b>	0.573 (1.544)	-0.293 (0.885)	0.728 (0.593)	-0.193 (0.375)	0.192 (0.169)	-0.157 (0.135)	0.050 (0.338)	-0.652** (0.328)	0.681 (0.557)	-0.832** (0.421)
<b>Respondent is head of household (1 = yes, 0 = no)</b>	0.350 (1.553)	-0.394 (0.898)	0.688*** (0.234)	0.168 (0.271)	0.070 (0.135)	-0.062 (0.089)	0.492 (0.302)	-0.026 (0.235)	0.813* (0.431)	-0.114 (0.286)
<b>Respondent age (= 1) * head of household (Interaction effect)</b>	1.400 (2.311)	1.912 (1.236)	-0.678 (0.809)	-0.032 (0.405)	-0.102 (0.228)	0.306** (0.151)	0.131 (0.465)	1.010** (0.422)	0.006 (0.769)	1.311** (0.509)
<b>Respondent education – secondary and higher (1 = yes, 0 = no)</b>	1.989 (1.832)	-0.207 (0.858)	0.298 (0.282)	0.003 (0.175)	0.373** (0.149)	0.045 (0.069)	-0.086 (0.312)	0.201 (0.203)	0.506 (0.521)	0.254 (0.250)
<b>Respondent gender (1 = male, 0 = female)</b>	0.356 (1.374)	-0.321 (0.806)	-0.564 (0.345)	-0.226 (0.198)	-0.027 (0.127)	0.035 (0.080)	-0.055 (0.258)	-0.236 (0.263)	0.082 (0.408)	-0.207 (0.296)
<b>Respondent agreeableness</b>	-1.274** (0.565)	-0.322 (0.236)	0.117 (0.208)	0.091 (0.113)	-0.018 (0.059)	0.006 (0.033)	-0.081 (0.145)	-0.076 (0.089)	-0.172 (0.255)	-0.084 (0.104)
<b>Respondent has agricultural background (1 = yes, 0 = no)</b>	-4.305*** (1.389)	-2.521* (1.282)	-1.086** (0.453)	-0.895*** (0.327)	-0.762*** (0.176)	-0.470*** (0.139)	-1.661*** (0.313)	-1.266*** (0.388)	-2.511*** (0.475)	-1.395*** (0.420)
<b>Interviewer gender (1 = male, 0 = female)</b>	2.511* (1.291)	-0.274 (0.552)	-0.163 (0.315)	-0.119 (0.141)	-0.112 (0.172)	0.023 (0.065)	-1.072*** (0.290)	0.810*** (0.198)	-0.568 (0.466)	0.769*** (0.225)
<b>Interviewer has completed degree (1 = yes, 0 = no)</b>	1.520 (1.188)	Omit. (0.731)	-0.017 (0.731)	Omit. (0.150)	-0.036 (0.150)	Omit. (0.358)	-0.470 (0.523)	Omit. (0.523)	-0.464 (0.014)	Omit. (0.007)
<b>Interviewer from suitable field of study (1 = economics/agriculture, 0 = other)</b>	-1.850* (1.024)	-0.096 (0.608)	0.347 (0.696)	-0.146 (0.164)	-0.284** (0.110)	-0.066 (0.065)	-0.671** (0.320)	0.401* (0.233)	-1.077** (0.447)	0.457* (0.255)
<b>Interviewer exam score (in %)</b>	0.096*** (0.035)	0.003 (0.021)	-0.008 (0.020)	-0.001 (0.004)	-0.009** (0.004)	0.001 (0.002)	-0.058*** (0.009)	0.009 (0.006)	-0.046*** (0.014)	0.010 (0.007)
<b>Interviewer household survey experience (1 = yes, 0 = no) (No experience is base)</b>	-0.882 (1.186)	-2.061 (1.308)	0.510 (0.486)	0.186 (0.191)	-0.235** (0.118)	0.045 (0.099)	-0.132 (0.315)	-0.753** (0.349)	-0.208 (0.465)	-1.069** (0.527)
<b>Interviewer TVSEP survey experience (1 = yes, 0 = no) (No experience is base)</b>	-1.696 (1.476)	-4.347*** (1.537)	0.103 (0.345)	0.132 (0.432)	0.227 (0.208)	-0.150 (0.133)	0.108 (0.329)	-0.384 (0.693)	0.041 (0.470)	-1.050 (0.774)
<b>Interviewer openness</b>	-1.119 (0.693)	0.780** (0.301)	-0.130 (0.225)	-0.034 (0.104)	0.001 (0.047)	-0.019 (0.038)	0.342*** (0.112)	-0.004 (0.102)	0.077 (0.195)	0.084 (0.125)
<b>Interviewer extraversion</b>	2.762*** (0.982)	0.329 (0.387)	0.074 (0.239)	-0.116 (0.127)	-0.178*** (0.065)	-0.036 (0.045)	-0.258* (0.141)	-0.084 (0.112)	0.194 (0.267)	-0.080 (0.135)

# Results OLS (II)

Table 3. Probit model results by country (cont.)

Independent variables	(1) Missing value		(2) Refusal value		(3) Outlier value		(4) Implausible value		(5) Total erroneous value	
	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam	Thailand	Vietnam
Congruent ethnicity (1 = yes, 0 = no)	Omit. (0.804)	0.038	Omit. (0.192)	-0.076	Omit. (0.063)	0.150** (0.063)	Omit. (0.194)	-0.227	Omit. (0.256)	-0.344
Log of interview duration (minutes)	-0.225 (1.410)	0.487 (0.822)	-1.544*** (0.765)	-0.618** (0.260)	-0.488*** (0.180)	-0.172 (0.109)	-0.142 (0.335)	-0.369 (0.353)	-0.481 (0.596)	-0.062 (0.396)
Morning interview (1 = yes, 0 = no)	-0.807 (0.983)	-0.010 (0.603)	0.085 (0.307)	-0.235 (0.153)	-0.011 (0.104)	0.119* (0.061)	0.469** (0.220)	-0.024 (0.161)	0.253 (0.349)	-0.039 (0.196)
Respondent participated in all waves (1 = yes, 0 = no)	-3.310** (1.573)	-0.763 (1.198)	-0.415 (0.471)	-0.232 (0.308)	0.168 (0.228)	-0.140 (0.137)	-0.001 (0.396)	0.071 (0.364)	-0.839 (0.517)	-0.012 (0.436)
Survey week	0.024 (0.431)	0.136 (0.224)	-0.089 (0.196)	-0.003 (0.075)	-0.137*** (0.047)	-0.003 (0.035)	-0.589*** (0.101)	-0.351*** (0.116)	-0.871*** (0.165)	-0.458*** (0.120)
Buriram province (1 = yes, 0 = no) (Nakhon Phanom is base)	-4.013* (2.363)	Omit. (0.468)	0.187 (0.468)	Omit. (0.164)	-0.253 (0.164)	Omit. (0.363)	0.181 (0.363)	Omit. (0.668)	-1.382** (0.668)	Omit.
Ubon province (1 = yes, 0 = no) (Nakhon Phanom is base)	-1.671 (2.329)	Omit. (0.258)	0.825*** (0.258)	Omit. (0.163)	-0.099 (0.163)	Omit. (0.356)	0.025 (0.356)	Omit. (0.681)	-0.517 (0.681)	Omit.
Ha Tinh province (1 = yes, 0 = no) (Dak Lak is base)	Omit. (0.723)	-1.154 (0.723)	Omit. (0.199)	-0.521*** (0.199)	Omit. (0.088)	-0.370*** (0.088)	Omit. (0.363)	-1.036*** (0.363)	Omit. (0.374)	-1.245*** (0.374)
Hue province (1 = yes, 0 = no) (Dak Lak is base)	Omit. (1.017)	1.236 (1.017)	Omit. (0.263)	-0.365 (0.263)	Omit. (0.108)	-0.265** (0.108)	Omit. (0.248)	-0.564** (0.248)	Omit. (0.284)	-0.437
Household size (persons)	0.539 (0.328)	-0.371** (0.154)	-0.119 (0.074)	-0.006 (0.032)	-0.084*** (0.029)	-0.051*** (0.017)	-0.329*** (0.064)	-0.198*** (0.048)	-0.125 (0.110)	-0.223*** (0.058)
Constant	5.897 (10.763)	2.118 (6.317)	10.308*** (3.385)	5.841** (2.310)	6.470*** (1.116)	2.581*** (0.723)	11.323*** (1.919)	9.575*** (2.908)	15.301*** (3.354)	9.043*** (3.051)
Observations	335	182	585	134	539	484	1,487	1,345	1,646	1,392
Pseudo R <sup>2</sup>	0.120	0.086	0.059	0.227	0.143	0.107	0.108	0.070	0.069	0.061

\* Significant at 10%; \*\* Significant at 5%; \*\*\* Significant at 1%.

Notes: Robust standard errors in parentheses

Source: Own calculation based on TVSEP survey 2017