

Sample Design

The sample, consisting of provinces, sub-districts (Thailand), communes (Vietnam), villages and households, follows a three-stage sampling design. In the first stage, provinces were chosen purposively based on official statistics, that met the following criteria:

- Low average per capita income,
- high dependence on agriculture,
- existence of special risk factors such as remoteness and peripheral location along the country's border,
- poor infrastructure,
- risky conditions for crop production (drought, flood, storms).

In Thailand, three provinces in the north-eastern part of the country, namely Buriram, Ubon Ratchathani and Nakhon Phanom, were selected. In Vietnam, the three provinces are in the Central Region, ranging from the north central coast, i.e., Ha Tinh, and Thua Thien Hue to the southern part of the central highlands, i.e., Dak Lak (see Figure 1).



Figure 1: Map of selected study provinces

Source: Hardeweg et al. (2013).

These six provinces have in common that they are peripheral to the border with Laos and/or Cambodia. They differ, to various degrees, with regards to their agro-ecological and agro-economic conditions, infrastructure, and their stage of development. Comparing the provinces allows drawing conclusions for regional development. The provinces can be taken as replicates to facilitate comparisons between the two countries.

The sampling procedure consists of a 3-stage cluster sampling design. The initial total sample size was 4,400 households in both countries. These were drawn from 440 villages and 220 sub-districts or communes.

In Thailand, the primary sampling unit (1st stage) are the sub-districts (Tambons), whereby urban sub-districts or towns were excluded. The Tambons were stratified in more densely populated, i.e., peri-urban, and less densely populated, i.e., rural sub-districts. In order to ensure proportional coverage of the rural population, systematic random sampling based on a list ordered by population density/size (PPS) was applied.

In the 2nd stage, two villages per sub district were sampled with a probability proportional to size of population from each of the sampled sub-districts. The selection probability (p) for village v in sub-district s and stratum r is given by equation (1):

$$p_{rsv} = \frac{a_r \cdot b \cdot m_{rsv}}{\sum_s m_{rs}} \quad (1)$$

whereby: a_r is the size of the primary sampling unit in stratum r , b is the number of villages sampled in each sub-district, m_{rs} and m_{rsv} is the population size of sub-district and village, respectively.

At the third stage, a fixed size sample “ c ” of 10 households has been selected systematically from a list of households ordered by household size with equal probability of selection (EPS). The selection probability for households as shown by equation (2) leads to a constant probability of selection for all households if a_r is determined proportionally to stratum size.

$$p_{rvh} = \frac{c}{m'_{rsv}} \cdot p_{rsv} = \frac{a_r \cdot b \cdot c}{\sum_s m_{rs}} \cdot \frac{m_{rsv}}{m'_{rsv}} \quad (2)$$

where m'_{rsv} is the number of households from the household listing frame.

An overview of the sample design is given in Table 1.

Table 1: Overview of sample design in Thailand

Stage	Sampling unit	Selection criterion	Sampling probability
Target population	Province	Purposive selection: Border provinces in North-Eastern Thailand, low income, significant dependence on agricultural income and assumed risky environment.	-
1 st	Sub-district	Provinces are constituted strata with approximately proportional sample size a_r , PPS systematic random sample with implicit stratification by population density	$\frac{a_r \cdot m_{rs}}{\sum_s m_{rs}}$
2 nd	Village	Simple random PPS sample of 2 villages from each sampled sub-district	$\frac{b \cdot m_{rsv}}{\sum_v m_{rsv}}$, b=2
3 rd	Household	Equal probability of selection' (EPS) systematic random sample with implicit stratification by household size	$\frac{c}{m'_{rsv}}$, c=10

Sampling frames were obtained from two databases maintained by the Department of Community Development. First, a village-level database (NRC2D), which provided the measure of size at the sub-district and village levels as of 2005 and second a household database (BMN) of 2006 which was used as a listing frame for rural households. Secondary data for sampling in Thailand were available at the village level, and population density and agro-ecological conditions were assumed to be sufficiently homogeneous within the province. Therefore, the sample is self-weighting by design. Furthermore, the sampling ratios for the three provinces are well above the socioeconomic surveys of the Thai National Statistical Office (NSO 2008). The NSO

cross-section survey for income, is carried out for rural and urban households, in all Thai provinces, including Bangkok. The NSO survey includes some 52,000 households, both rural and urban (IHSN, 2022). This equates to an average of about 675 households per province for both urban and rural households. Since the overall ratio of urban to rural households is about 1.08, the TVSEP sample size is more than double that of the NSO survey. In the three TVSEP provinces, the sample size for rural households ranges from 400 in Nakhon Phanom to 960 in Ubon Ratchathani.

The average sample size per province of 733 households, both rural and urban, is less than one half of the average provincial sample size of TVSEP when comparing rural households between TVSEP and the NSO. The same is true for the Thai Townsend project (Binford et al., 2004) which reports a sample of 48 villages per province (TTTP, 2022). The project spans four provinces, one of them is the TVSEP province Buriram, where TVSEP has 84 villages in the sample.

In contrast to the sample from Thailand, the three Vietnamese provinces in the panel are quite diverse in terms of natural conditions. The most central province is Thua Thien Hue, with its capital Hue City. It was at the time of panel implementation the most developed among the three provinces. It shares a 128 km coastline with the South China Sea to the east and is bordering Laos in the west. The same is true for Ha Tinh, the most northern of the three TVSEP provinces whose coastline with the South China Sea spans 137 km. Ha Tinh has a long mountain range border with Laos and in 2007, it was the least developed among the three TVSEP provinces. The third province, Dak Lak, is the most southern of the TVSEP provinces, some 300 km north of Ho Chi Minh City. It is landlocked and located in the Central Highlands, and is distinct due to its high population of ethnic minorities, especially the Ede group, an Austronesian ethnicity. Dak Lak is known for its coffee plantations, dating back to the French colonial time.

To capture the heterogeneity in the sample design, strata were defined according to geographic and ecological conditions. In the two provinces of the northern part of the Central Highlands these were coastline, lowlands and uplands, while in Dak Lak only the two latter zones applied. In order to accommodate the low population density in the upland areas of these provinces, an absolute minimum sample size was fixed at 160 households per strata. For the first stage sampling units, i.e., the commune level (equivalent to sub-district in Thailand), no measure of size was available at the time of

sampling. Instead, the population share of the respective district was used for the weighting of commune selection. The selection probability of a commune is thus defined by the first fraction on the right side of equation (3)¹.

$$p_{rdsvh} = \frac{a_r \cdot m_{rd}^*}{N_{rd} \sum_d m_{rd}^*} \cdot \frac{2 \cdot m_{rdsv}^*}{\sum_v m_{rdsv}^*} \cdot \frac{10}{m_{rdsv}} \quad (3)$$

At the second stage, villages were sampled with probability proportional to size based on population m^*_{rdsv} (see equation (3)). The third stage was again a systematic random sample drawn with equal probability from the village household lists ordered by household size. This is the recommended strategy for the last stage in clustered sampling in order to capture a maximum of variation within the cluster. Data for the local administrative units and household sample frames were taken from the Agricultural and Rural Census from 2006, which covered all rural households and was conducted by the Vietnam General Statistical Office.

The stratification applied in Vietnam included the specification of agro-ecological zones as analytical domains. Setting a lower bound to the absolute stratum sample size for the sparsely inhabited strata leads, to varying ranges of the selection probabilities of a household ranging from 0.147% to 5.85% (Table 2). As a consequence, analysis of the survey data that aim at generating information about the provincial situation, e.g., provincial vulnerability profiles, require a weighting procedure. The sampling ratios for the provinces are almost twice those obtained in the Vietnam Household Living Standard Survey (VHLSS). The VHLSS usually does not implement all of its modules in the survey instrument with the entire sample. For example, the most recent survey collected expenditure data from 30,000 rural and urban households in 61 provinces, i.e., around 490 per province (World Bank Microdata Library, 2022a). For other modules, 45,000 households were considered, i.e., on average around 740 households per province. As TVSEP is focused on rural households, its sample size is roughly twice that of the VHLSS in the three TVSEP provinces (World Bank Microdata Library, 2022b).

¹ For explanation of symbols see equation 2

Table 2: Basic data of the target population and TVSEP household sample in Vietnam, 2005

Province/Strata	Rural Population	Population density (Pop./km ²)	Pop. share (%) total province	Sample allocation		Selection probability
				absolute	(%)	range
						(%)
Dak Lak	1,335,193	102	41	760	35	
<i>Rice plain</i>	452,982	64	34	260	34	0.165 - 0.603
<i>Mountainous area</i>	882,211	145	66	500	66	0.164 - 1.766
Thua Thien – Hue	788,763	156	24	720	33	
<i>Coastal area</i>	376,693	322	48	240	33	0.224 - 1.074
<i>Rice Plain</i>	357,612	179	45	240	33	0.175 - 0.575
<i>Mountainous area</i>	54,458	29	7	240	33	0.624 - 5.85
Ha Tinh	1,147,693	191	35	720	33	
<i>Coastal area</i>	567,609	246	49	360	50	0.196 - 0.783
<i>Rice Plain</i>	338,781	489	30	200	28	0.171 - 0.536
<i>Mountainous area</i>	241,304	80	21	160	22	0.147 - 0.724
Total:	3,271,649			2200		

Source: Provincial Statistical Year Books 2005, General Statistics Office, Hanoi;
own calculations.

References

- Binford, M. W., Lee, T. J., & Townsend, R. M. (2004). Sampling design for an integrated socioeconomic and ecological survey by using satellite remote sensing and ordination. *Proceedings of the National Academy of Sciences*, *101*(31), 11517-11522.
- Hardeweg, B., Klasen, S., & Waibel, H. (2013). Establishing a database for vulnerability assessment. In *Vulnerability to Poverty* (pp. 50-79). Palgrave Macmillan, London.
- The Townsend Thai Project (2022, September 14). Baseline Survey (“The Big Survey”). Retrieved from: <http://townsend-thai.mit.edu/data/baseline-survey.shtml>. Accessed on: 14.09.2022.
- World Bank Microdata Library (2022a, September 14). Household Living Standards Survey 2002. Vietnam, 2002. Retrieved from: <https://microdata.worldbank.org/index.php/catalog/2306>. Accessed on: 14.09.2022.
- World Bank Microdata Library (2022b, September 14) Household Living Standards Survey 2004. Vietnam, 2004. Retrieved from: <https://microdata.worldbank.org/index.php/catalog/2370>. Accessed on: 14.09.2022.