



Tropentag 2007
University of Kassel-Witzenhausen and
University of Göttingen, October 9-11, 2007
Conference on International Agricultural Research for Development

Sampling for vulnerability to poverty: cost effectiveness versus precision

Bernd Hardeweg^a, Suwanna Praneetvatakul^b, Tung Phung Duc^c and Hermann Waibel^a,

a Leibniz University Hannover, Faculty of Economics and Management, Königsworther Platz 1, 30419 Hannover, Germany, email: hardeweg@ifgb.uni-hannover.de

b Kasetsart University, Dept. of Agricultural and Resource Economics, Faculty of Economics, Bangkok, Thailand

c Institute of Policy and Strategy for Agriculture and Rural Development, Center for Agricultural Policy, Hanoi, Vietnam

Abstract

This paper presents the procedure of measuring and sampling for vulnerability to poverty by means of establishing a database in Thailand and Vietnam. It reports about work in progress and introduces the theoretical and practical issues of data collection and sampling in the context of a medium-size research project in the field of economics. Possibilities and constraints for data collection that exist under the conditions of limited secondary information and the usual budget constraints are shown. The paper addresses two issues. First questions arising with the design of a questionnaire that combines various fields of economics and aims to measure respondents' notion of risks, shocks and vulnerability are discussed. Second sample design and sampling procedure applied to a sample of 2200 households in two countries each are shown. Results show that it is possible to effectively conduct a vulnerability survey and apply a statistically satisfactory sampling method given cost constraints and availability of secondary data. Two major challenges can be mentioned. One is to capture the socio-cultural differences that exist in terms of perception on risk, shocks and vulnerability and the second is to adequately reflect the heterogeneity of the study population and matching this with available secondary information.

Introduction

Until recently development institutions and economists have measured the well-being of poor people and households in a population using cross-sectional data on income and/or consumption expenditures mostly over a short period of time. The focus was on static poverty measures generally expressed by indicators such as the headcount ratio and the poverty gap. Recent research however has established that for assessing the well-being of poor households and for devising effective strategies of poverty reduction in developing countries measures are needed that take into account the dynamic nature of poverty phenomena (e.g. LIGON & SCHECHTER, 2003). Such measures must have a predictive capacity to recognize the fact that a household who is not poor today might yet be vulnerable to falling into poverty as a result of covariate (e.g. natural disasters, financial crises, epidemics) and idiosyncratic (e.g. death, injury, unemployment) shocks. As vulnerability is a function of (a) the occurrence and severity of various idiosyncratic and covariate past shocks and future risks and (b) the ability of a household to cope through ex post and ex ante insurance mechanisms the data collected under conventional poverty assessment regimes usually do not meet the requirements to measure vulnerability. What is needed instead are panel data over a longer period of time and a well designed questionnaire that facilitates the

quantification of the components of vulnerability. Also a large sample of households who either are below the poverty line or who may not be poor today but who face the risk of being poor in the future is needed. Collection of such information is a challenge as usually insufficient prior information is available.

The research presented in this paper is part of a larger DFG research project “Impact of shocks on the vulnerability to poverty: consequences for development of emerging Southeast Asian economies” involving the Universities of Hannover, Göttingen and Frankfurt. A survey among a total of 4400 households in six provinces in Thailand and Vietnam has been conducted in 2007.

Objectives

This paper reports the procedure of measuring and sampling for vulnerability to poverty by means of establishing a database that allows advancing the concept of vulnerability in two Southeast-Asian countries. The paper illustrates the possibilities and constraints for data collection that exist under the conditions of limited secondary information and the usual budget constraints.

Methodology and Data

Concept of the Questionnaire

To operationalize measurement of vulnerability to poverty a questionnaire has been designed that aims at capturing the various components of vulnerability. Different areas of economics including agricultural economics, economics of poverty dynamics, financial institutions and economic geography have jointly developed the questionnaire. The challenge when assessing vulnerability is to incorporate dynamic aspects which allow to develop empirical models with variables that can explain the conditions that make a household vulnerable. On the one hand this requires to include the conventional parameters of household surveys like the social environment of the household, its resource endowments, household income and consumption expenditures and in addition include questions that facilitate the quantification of past shocks experienced by the household and perceived future risks.

Two questionnaire forms were developed. The first questionnaire with a size of 3 pages and 95 variables targets the village head and collects information on village infrastructure and economy and social problems from one for the household interview. The comprehensive household questionnaire covers the demographic, economic and social situation of the household and perception of risks on a total of 61 pages containing 420 variables. Multiple data rows per household (e.g. for data on each household member) lead to an average of 910 data items collected per household.

In Table 1 the main criteria and indicators for the vulnerability measurement in the household questionnaire are summarized.

The questionnaire combines the elaboration of facts based on the memory of the respondent with his or her perceptions on future events influenced by his or her experience and judgements of the future. Thus a trade-off between the degree of detail on measuring the household's current level of well-being and the elaboration of his/her experience with shock events and future risks had to be taken into account. The data of the first survey allow conducting a cross sectional ex post impact assessment of shocks experience by the household. In addition ex ante assessment of the household's well-being is facilitated by drawing on variables that express the subjective probabilities and expected consequences of risky events.

Table 1: Criteria and indicators for measurement of vulnerability in household questionnaire

Criteria	Indicators
Resource Endowment	Land, Buildings, Farm and Household Assets, Knowledge, Health, Access to common property resources, Social capital
Household Organisation	Ethnicity, Size, Dependency, Gender, In- and Out-migration
Sources of Income	Farm, Off-Farm, Non-Farm, Transfers
Household Consumption	Food, Non-Food, Education, Health, Transportation
Sources of Vulnerability	Past shocks (idiosyncratic, covariate), perceived risks (magnitude and probability of event)
Risk management	Income diversification, Contracting, Saving, Borrowing, Lending, Investments including Social Capital

Sample Design

In line with the overall objective of the project, the survey targets rural households either poor or those who are at risk of falling into poverty from two different Southeast Asian economies. For this purpose, three peripheral provinces in Northeast Thailand and three provinces from the North Central Coast and Central Highlands in Vietnam were deliberately selected for their peripheral location along a border to their common neighbour Laos or Cambodia and a certain degree of variation of agro-ecological conditions.

The sampling design aimed to obtain a representative sample of the target population of rural and peri-urban households while containing survey cost by using an appropriate cluster sampling technique. The procedure was based on the guidelines of the UN Department of Economic and Social Affairs (UN 2005) for household survey samples. The sampling procedure consists of a 3-stage cluster sampling design. The budget available has limited the total sample size to 2,200 households per country. The ultimate cluster size of 10 households in a village was chosen based on organizational aspects of the survey, i.e. the size of the survey team but is also in line with recommendations and prior information for Vietnam that homogeneity within villages is high (PETTERSSON, 2003).

Thailand household sample

As secondary data for sampling for Thailand were available down to the village level and population density and agro-ecological conditions were assumed to be sufficiently homogeneous, sample design for Thailand could be kept simple and aimed at obtaining a self-weighting sample. However, in order to contain travel cost, geographic dispersion of sample villages was controlled by selecting 110 sub-districts in the first stage before sampling two villages from each sampled sub-district in the second stage. This pattern enabled survey teams to finish interviews in two villages per day and at the same time ensures sufficient geographic coverage of the sample.

The details of the sampling procedure for Thailand are shown in Table 2. In the first stage, provinces were treated as constituted strata with approximately proportional allocation of primary sampling units (PSU: sub-districts). We ignored the district level because these are administrative units with no expected impact on vulnerability differences. Sub-districts were selected with probability proportional to population size (PPS). In order to ensure proportional coverage of densely (peri-urban) and less densely populated (rural) areas systematic random sampling based on a list ordered by population density was applied. The measure of size was the number of households as of 2005. At the 2nd stage two villages were sampled with probability proportional

to size from each of the sampled sub-districts,. The selection probability p_{rsv} for village v in sub-district s and stratum r is given by equation (1).

$$(1) \quad p_{rsv} = \frac{a_r \cdot m_{rs}}{\sum_s m_{rs}} \cdot \frac{b \cdot m_{rsv}}{\sum_v m_{rsv}} = \frac{a_r \cdot b \cdot m_v}{\sum_s m_{rs}}, \quad \text{because } \sum_v m_{rsv} = m_{rs},$$

where a_r is the PSU sample size in stratum r , b is the number of villages sampled in each sub-district, m_{rs} and m_{rsv} are the measures of size by sub-district and village, respectively.

Table 2: 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	EPS systematic random sample with implicit stratification by household size	$\frac{c}{m'_{rsv}}, c=10$

At the third stage a fixed size sample 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 given by equation (2) leads to a constant probability of selection for all households if a_r is determined proportionally to stratum size.

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

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

In order to ensure smooth continuation of field operations in case households were found ineligible, replacement samples have been added. Sampling frames were obtained from two databases maintained by the Department of Community Development, Ministry of the Interior. The village-level database (NRC2D) provided the measure of size at the sub-district and village levels as of 2005. The household database (BMN) for the three provinces of 2006 became available at a later stage and was used as a listing frame for rural households including household size.

Vietnam household sample

The provinces purposively selected for the survey in Vietnam are geographically more diverse than those in Thailand. While Dak Lak province is part of the landlocked Central Highland, Thua Thien-Hue and Ha Tinh provinces extend from the coast to the mountainous border to Laos. In order to take into account this heterogeneity, strata for the first stage were defined as agro-

ecological zones within the three provinces. As reported in Table 5, these strata differ significantly in size and a proportional allocation of 2,200 samples would lead to insufficient absolute sample sizes in some of the strata. For this reason, different sampling rates were applied by stratum such that the ultimate absolute sample size was fixed at a minimum of 160 households, in order to allow meaningful inferences also for these entities. As for Thailand, a first stage sample of sub-district preceded the village sample to reduce travel cost between villages. For the first stage sampling units (sub-districts or communes) no measure of size was available at the time of sampling. Instead the population share of the respective district d was used for weighting sub-district selection. The selection probability of a sub-district is thus defined by the first fraction on the right and side of equation (3).

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

At the second stage, villages were sampled with probability proportional to size based on population m_{rdsv}^* as given by the second part in equation (3). The third stage was again a systematic random sample with equal probability from household lists ordered by household size. This is the recommended strategy for the last stage in cluster 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 2006, which covers all rural households and has been conducted by the Vietnam General Statistical Office.

Results

Interviews

The conduct of the survey among 4400 households in Thailand and Vietnam revealed that it is possible to engage respondent in a discussion of past shocks. Respondents do have their own notion of shocks. Also respondents are willing share experience about their life and subsequently find it easier to assess future risks. The change in topics from factual information to events and perceptions enhances the cooperation with the interviewer. On average respondents reported 1.4 shocks during the past five years and provide assessment of 19 proposed risky events out of which they expect about 10 (Thailand) and 6 (Vietnam) to occur in the coming 5 years albeit at highly varying degrees of frequency. However interviewers are challenged to effectively navigate through the questionnaire making good introductions of the different sections of the form in order to improve interview atmosphere. Also such interviews are time consuming usually between two and three hours and therefore require good incentives for the respondents and well trained interviewers.

Sampling

To generate a vulnerability data base that allows improving our understanding of the causes and mechanisms of vulnerability to poverty is a challenge. Given the usual resource limitations of a medium-size research project such data may not be representative for a larger administrative area usually required by administrators and policy makers when implementing poverty reduction programs. However it may be possible to draw conclusions for smaller geographic units such as the provinces selected for this research. The major problem is the limitation of sufficient prior information at the level of the main enumeration unit, which is the rural village. The application of a multi-stage sampling design in connection with sufficiently up to date secondary data at high spatial resolution reduces the amount of data required for random selection at the household level, which simplifies data handling and also makes access to existing listing frames from statistical agencies easier. The experience of the survey has shown that such frames of sufficiently recent date exist in both countries, and can be made available for use in a research

project. Even though listing frames were of comparatively recent date, the survey has revealed errors and gaps such as households that had moved away or been disbanded in a few places, however. As a result of the simple sample design selection probability for households in Thailand was fairly uniform in a range from 0.358 to 0.361 per cent (Table 3), and reflects the homogeneity of the survey area.

Table 3: Basic data for target population and household sample in Thailand

	Rural population	No. of households	Share of provincial strata	Number of sub-districts (PSU)	Sample size (households)	Probability of selection for a household
Buriram	956,497	228,823	37.6%	184	820	0.3584%
Ubon Ratchathani	1,142,219	271,213	44.6%	215	980	0.3613%
Nakhon Phanom	444,562	108,662	17.9%	95	400	0.3681%
Total/Average	2,543,278	608,698	100.0%	494	2,200	0.3614%

Data: Village database 2005, Community Development Department, Ministry of the Interior

The situation in Vietnam was somewhat different from Thailand. Here we were faced with a high degree of heterogeneity of agro-ecological and economic conditions with implications for the composition of income portfolios of the rural population (see Table 4). The variation of the agricultural income share is especially pronounced in provinces of Thua Thien-Hue and Ha Tinh. The densely inhabited coastal areas, the inland rice plains with medium population density and the sparsely populated mountains provide are factors that need to be considered in the stratification.

Table 4: Selected income indicators for the rural population in the survey provinces

	Average household income from agriculture (1,000 VND)		Share of income from agriculture	
	Mean	CV	Mean	CV
Dak Lak	10,948	73%	78%	37%
Thua Thien - Hue	4,880	96%	35%	82%
Ha Tinh	5,613	96%	55%	50%

Source: Vietnam Household Living Standard Survey 2004

Table 5: Basic data for the target population and household sample in Vietnam

	Rural Population	Population density (1/km ²)	Population share of total of province	Sample allocation		selection probability range %
				absolute	%	
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.

Hence the stratification applied in Vietnam had to be adjusted to include agro-ecological zones. Setting a lower bound to absolute stratum sample size for the sparsely inhabited strata leads, however, to varying selection probabilities for the Vietnam sample in the range from 0.147% to 5.85% (Table 5) As a consequence, while the Thailand sample is approximately self-weighting, analysis of survey data from Vietnam will be complicated by the need for applying a weighting procedure for generating provincial vulnerability profiles for example.

Discussion

The survey conducted among rural households in 220 villages of Northeastern Thailand and another 220 villages in three provinces of the North Central Coast and Central Highland in Vietnam has been successfully completed and shown that it is possible to obtain a comprehensive data set for vulnerability assessment in a single interview session. A preliminary assessment of responses to questions on risk perceptions shows pronounced differences among the two countries included in the survey and warrant further in-depth analysis.

In the planning phase, costs and organizational advantages of a multi-stage sampling procedure had to be weighed against additional complexity during analysis as a result of the complex sampling design and unequal selection probabilities. In view of the heterogeneity of study sites in Vietnam, the original plan of an analogous procedure in both countries was abandoned in favour of stratification according to agro-ecological zones including oversampling of small strata at the cost of lower overall sample reliability. Although the quality of the sample can be judged only after estimation of sampling error and design effect (i.e. the increase of sampling error due to deviation from simple random sampling), a flexible adaptation of the sampling procedure to organizational requirements but also the availability of prior information was indispensable for a successful execution of the survey.

The total cost of the survey, including preparation, training of interviewers, survey operations, data entry and data cleaning amount to about US\$ 48 per household questionnaire. We consider this a reasonable amount for an average of 900 data items collected by a rather complex questionnaire. Yet the costs are significantly below rates that have been quoted e.g. for the Vietnam Housing and Living Standards Survey of US\$ 80 – 120 per questionnaire.

References

- LIGON, ETHAN & LAURA SCHECHTER (2003): Measuring Vulnerability. *The Economic Journal*, **113 (March)**(486), p. pp. 95–102.
- PETTERSSON, HANS (2003): Mission Report: Recommendations Regarding the Sample Design of the Vietnam Household Living Standard Survey. Unpublished mission report, Statistics Sweden.
- UNITED NATIONS - DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS (2005): Designing Household Survey Samples: Practical Guidelines. Studies in Methods // United Nations, Department of Economic and Social Affairs, Statistics Division Series F ; 98; United Nations, New York.