The Importance of Being Surveyed: The Representativeness and Impact of the Vietnam Household Living Standards Surveys

By HENRIK HANSEN AND TRUNG DANG LE*

January 2013

Using census data from 1999 and 2009 we test the representativeness of the Vietnam Household Living Standards Survey (VHLSS) and its impact on households' possessions of durable goods, housing conditions and educational attainment. We find that households residing in VHLSS communes were on average better off than households living in other communes both in 1999 and 2009. We also find that the shares of households with better housing conditions and educational attainment increased more in the VHLSS communes than other communes from 1999 to 2009, indicating a causal effect on household wellbeing of communes being included in the VHLSS sample.

* Hansen: Professor, Department of Economics, University of Copenhagen, Øster Farimagsgade 5, bygning 26, DK-1953 København K, Denmark (email: henrik.hansen@econ.ku.dk); Le: PhD Student, Department of Food and Resource Economics, University of Copenhagen, Rolighedsvej 25, DK-1958 Frederiksberg C, Denmark (email: trung@foi.ku.dk). We would like to thank John Rand and seminar participants at Department of Food and Resource Economics, University of Copenhagen for helpful discussions and comments.

I. Introduction

The importance of the Living Standard Measurement Surveys (LSMS) is hard to overstate. The standard globally comparative measure of headcount poverty is the percentage of the population living below a 'dollar-a-day' poverty line (now actually $1.25). This important measure is estimated and revised using the LSMS and similar national household surveys. The dollar-a-day measure is known to most people and enshrined in the Millennium Development Goals; the first goal is to cut in half global headcount poverty, measured against this dollar-a-day poverty line.

Chen and Ravallion (2010) used 454 LSMS-type income and expenditure surveys from 97 developing countries to estimate the global poverty trends from 1980 to present. Almost all of these surveys are readily available for researchers and the public in general at the World Bank homepage and there is a designated search tool: http://iresearch.worldbank.org/lsms/-lsmssurveyFinder.htm. Researchers have spent thousands of man-years analyzing the LSMS data. The World Bank homepage lists more than 130 papers in the LSMS Working Paper Series

and refers to numerous papers using LSMS-data and a search for the phrase 'Living Standard Measurement Survey', using Scholar Google, results in more than 1,000 hits, not including citations.

An important feature of LSMS and related surveys is that they are viewed as nationally representative hence authoritative. Most assessments of trends in poverty or the effects of shocks on poverty, globally or in individual countries, are based on these surveys.

More importantly, the surveys have a direct impact on national policies. First of all, the geographical distribution of poverty, guiding the targeting of pro-poor policies and programs, is often estimated from LSMS-data and the World Bank has pushed the geographical targeting to the boundary with its endeavor into local area estimation of poverty using the LSMS data in connection with population census data (Elbers, Lanjouw, and Lanjouw, 2003; Bedi, Coudouel, and Simler, 2007). Second, most countries now have repeated surveys, forming a sequence of national poverty measures, whereby the success or failure of poverty programs and policies is evaluated by the change over time in the estimated poverty levels and ratios. Finally, World Bank country offices and national researchers and policy makers typically regard the LSMS data as the best and most reliable information they can obtain. Sometimes they even forget that the data is only a small general purpose survey thereby confusing estimates with population information. If the data is truly representative this is not a big problem. However, we don't really know if the surveys are representative. Much effort and expertise has gone into the sampling frames, the sampling design, and the practical implementation of the surveys but, as always, the devil is in the detail.

One detail that seems to have gone unnoticed is that in some countries the survey areas have been the same over a prolonged period of time. Vietnam is our case in point. The statistical office in Vietnam (The General Statistical Office, GSO) took responsibility for the living standard surveys from 2002 onwards and renamed the surveys to the Vietnam Household Living Standard Surveys (VHLSS). Assisted by a Swedish survey statistician and World Bank staff, the GSO has conducted four survey rounds: VHLSS 2002, 2004, 2006, and 2008. The sampling frame was based on a population census from 1999 and the point to note is that, although the households in the four survey rounds were renewed - apart from a deliberate rotating panel section - the selected communes were kept fixed for all four rounds. Hence, at the commune level we have a decade of VHLSS and non-VHLSS participants.

We use this fact, in combination with two population censuses (1999 and 2009), to test if the Vietnam Household Living Standard Surveys (VHLSS) can be considered to be representative. The idea is to test if the average household endowment of certain durable goods, housing conditions and educational attainment, recorded in the population census, are equal for the VHLSS and non-VHLSS communes in the successive census years. If the VHLSSs are representative there should be no differences in the predictions of average household attributes. If they differ the representativeness can be questioned and, most importantly, if they diverge over time there is a possibility of a VHLSS effect. This would question the level and trend in national poverty estimates, as aggregated up from the VHLSS.

We find that households residing in VHLSS communes were on average better off in 1999 and are still better off in 2009 than the households living in non-VHLSS communes as durable goods such as televisions, motorbikes, phones, refrigerators and even computers are more widespread in the VHLSS-communes. Moreover, the housing is better because significantly more households are connected to the national power grid, live in permanent houses, and have access to clean drinking water and a hygienic toilet (a loo or a water closet). The differences between the two groups of households are larger for the rural areas than for the urban areas, indicating a possibly important bias in the rural poverty estimates for Vietnam.

For the quality of housing, and education attainment we also find that households in the VHLSS communes have, on average, fared better than the households in non-VHLSS communes in the 2000s. In contrast, we find the reverse for connection to the power grid and possession of a television. Nonetheless, the divergent path in the quality of housing and educational attainment calls into question if the poverty estimates, based on the VHLSS, over-state the poverty reduction in Vietnam in the 2000s.

The paper is organized as follows. Section II briefly introduces the Vietnam Household Living Standards Survey. Section III describes the census data and our outcome variables. In Section IV we present our method of testing the representativeness and causal effects of the VHLSS. Section V gives our empirical results, and Section VI concludes.

II. The Vietnam Household Living Standard Surveys

The Living Standards Measurement Study survey was brought to Vietnam in the early 1990s to provide reliable data for monitoring of living standards, evidence-based policy design, and evaluation of policies and programs. The survey, officially called ‘Vietnam Living Standards

Survey' (VLSS), was designed to be representative at the national level. The first two rounds were conducted in 1992/1993 and 1997/1998 by the State Planning Committee and the General Statistical Office (GSO). The United Nations Development Programme (UNDP) and the Swedish International Development Agency (SIDA) gave financial support while the World Bank provided technical assistance and staff to work with the survey team. The VLSS 1992/1993 included 4,800 households in 150 communes, while the 1997/1998 VLSS increased the sample to 6,000 households. Both surveys were stratified along administrative (geographical) regions but with probability proportional to size sampling to make the samples self-weighting.

Inheriting the technical capacity from conducting the two rounds of the VLSS survey, the GSO started the second phase of the household surveys in 2002. The surveys were officially renamed to the 'Vietnam Household Living Standards Surveys' (VHLSS) and it was decided to carry out a survey round every second year. The VHLSS surveys were different from the VLSS in many ways. An important change was the sample size, in terms of both the number of households as well as the number of communes participating in the surveys. In each round of the VHLSS about 3,000 communes were selected to participate, accounting for nearly 30 percent of all communes in Vietnam. In each chosen commune 3 households were interviewed creating a total household sample size of 9,000.[1]

The VHLSS data are considered to be of high quality and it is widely used by both national and international research communities (for recent published analyses using the data see *e.g.*, Mont and Nguyen, 2011; Imai, Gaiha, and Kang, 2011; Oostendorp, Tran, and Nguyen, 2009). Within the administrative and political community, it is generally agreed that the VHLSSs provide legitimate nationally representative household data for Vietnam. Survey estimates, using sampling weights, are routinely referred to as the national levels. In particular, poverty headcount rates estimated from the VHLSS data are published as the official poverty rates and it is always the major, if not sole, source of data for poverty assessments by the World Bank and the Government of Vietnam (see World Bank, 2003; VASS, 2007; VASS, 2011).

The VHLSS data have also been used as baseline data for several research projects. Nakata, Sawada, and Tanaka (2009) used the 2006 VHLSS as their base and conducted a resurvey of 2,000 households in 4 provinces in 2008. More recently, Benjamin et al. (2012) used the 2002 VHLSS as "an excellent baseline survey" for their evaluation of a redistributing land reform in

[1] Except for the VHLSS 2002, for which the sample size is 30,000 households.

the Central Highlands in 2002. To construct a "before and after" data set, they revisited 1,250 households in 50 communes of the 2002 VHLSS in 2008. The VHLSS sample is also used to create data for long-term follow-up programs that result in many revisits. The Vietnam Access to Resources Household Survey (VARHS) is an example. The survey was initiated in 2006, it was designed to follow about 2,300 households in 466 communes of the 2002 and 2004 VHLSSs (Barslund and Tarp, 2008; Markussen, Tarp, and Van Den Broeck, 2011). The VARHS has been repeated biennially with the latest round being conducted in 2010.

In this study we make use of the fact that the four VHLSSs between 2002 and 2008 share the same sampling design. The VHLSS is documented in Phung and Nguyen (2007), and they introduce the sample design as follows:

"The series of VHLSS from 2002 to 2010 rely on a master sample for sample selection. The master sample is a random sample of the 1999 Population Census enumeration areas. From this sample of enumeration areas, multiple samples of households can be selected for different surveys or for the different years of a rotating panel survey such as the VHLSS. The master sample used in the VHLSS is a two-stage area sample where communes are selected in the first stage and 3 enumeration areas (EA) per commune are selected in the second stage. The communes were stratified on province and urban/rural and the sample was allocated over strata proportional to the square root of the number of households. Both communes and EAs are selected with probability proportional to size (PPS), with the size being the number of households according to Population Census 1999." (Phung and Nguyen, 2007, p. 242).

To be precise, the stratification, based on province and an urban/rural partition, resulted in 122 strata because all 61 provinces, existing in 1999, have both urban and rural areas. The first step of the sampling procedure selected some 3,000 communes, serving as Primary Sampling Units (PSU). The stratification and geographical placement of the selected communes is shown in Figure 1. Within each stratum, in which the number of communes to select was fixed, communes, EAs, and finally individual households were selected by systematic sampling using the "every *k*th rule" at each stage to ensure broad spatial coverage.

As the VHLSS is a general purpose survey and the objects being measured have little, if any, systematic relation to the geographically ordered sampling lists within strata, the systematic sampling is expected to be essentially equivalent to simple random sampling, as shown in

Cochran (1977). This is why the VHLSSs are generally considered as being based on stratified random samples.

For our study, the sampling of households within communes is inconsequential. The important design feature is that the selected communes were the same in all four rounds of the survey. Thus, although villages within each commune changed from one round to another, apart from the subset purposely kept forming a panel data set, all the communes that participated in the 2002 VHLSS are repeated thought-out from 2002 to 2008. And, as seen from the description of the sampling design, this was known from the planning of the first survey round in 2000-2001.

We will use the survey design information for our tests of the representativeness and impact of the VHLSSs but we will not make direct use of any of the collected VHLSS data.

III. The Census Data

Our main data are subsets of the two rounds of the Population and Housing Census conducted in 1999 and 2009 (henceforth the 1999 Census and 2009 Census). The 1999 Census is a status of the Vietnamese population as of April 1st 1999. As already noted, this census was used as the sampling frame for the VHLSSs from 2002 to 2008. We have a random sample covering 33 percent of all the households of the 1999 Census. This is by far the largest sample made available to the public by the GSO and it has been used in studies estimating poverty mappings for Vietnam (Baulch et al. 2004; Hansen, Pham, and Vu 2007; Nguyen, Tran, and Van Der Weide 2010) and also for a study of birth year preferences (Do and Phung, 2010)

The 2009 Census was carried out in April 2009, exactly 10 years after the 1999 Census. Two questionnaires were used to collect the data. A short questionnaire was administered to all households to obtain basic demographic and housing information while a long questionnaire with additional questions giving details about demography, housing conditions, and possession of durable goods and assets was sent to a 15 percent sample of households.

CPHCSC (2010) describes the sampling design for the long questionnaire. In the first step, the GSO determined the sample size allocation for each district in Vietnam. There is an over-sampling of small districts to ensure the sample is representative at the district level. In the second step, within districts, participating households were randomly chosen. As a result, 10,896 communes were selected, accounting for more than 98 percent of all communes in Vietnam. We have the responses of the long questionnaire covering the 15 percent of the population in 2009

and also the sampling weights such that we are able to estimate population means and proportions across subsets of the 2009 population.

The administrative structure in Vietnam has changed over the ten-year period from 1999 to 2009 such that the Census data are not directly comparable at the lower administrative levels. In 1999 there were 61 provinces and 10,475 communes. In 2009, there were 11,112 communes in 63 provinces. Changes at the provincial level happened in 2003, 2004 and 2008. For example, in November 2003, the Dien Bien province was formed from part of Lai Chau; in January 2004, the Dak Nong province was formed from Dak Lak and Hau Giang was formed from Can Tho and in 2008 Ha Tay was merged with Hanoi. Furthermore, changes at the district and commune levels happened even more often. Common patterns of change were upgrading from a rural commune to an urban ward; formation of new communes, usually from splitting or merging; and renaming of communes.

In addition to the changes in the administrative structure, the commune coding system changed from 1999 to 2009 but an official mapping of the two systems does not exist. The 1999 Census follows a coding system using 7-digit numbers to identify communes. The first three digits identify provinces, the next two digits identify districts within the province and the last two digits identify communes within the district. The coding system for the 2009 Census is a 1-to-N sequence. For instance, the 63 provinces can be identified by numbers ranging from 1 to 96; the district identifier is coded by numbers from 1 to 973; and communes are identified by numbers from 1 to 32,248.

We link the two coding systems by a mapping based on commune names. Technically, it is not a good solution to use a string variable for combining data since the value is very sensitive to typos, and the many diacritical signs in the Vietnamese spelling make matters worse. Fortunately, we have been successful in developing algorithms automating the combination of data for most of the communes. Out of 10,475 communes in 1999 we identify and link 10,211 communes using a reproducible protocol. The remaining communes had to be linked manually either because their names had changed completely or because the communes did not link one-to-one as a consequence of commune splitting or merging. The manual linking was done by referring to the legal documents that instructed the changes. A decision about the links in the cases other than one-to-one mapping had to be made to establish a complete link between the 1999 and 2009 censuses. We decided to combine smaller communes with the same source to

merge with the original commune (the source). By this process we finally obtain a linked data set of 10,277 communes in 1999 and 2009.

The census questionnaires are directly comparable except for the coverage of durable goods, which is larger in the 2009 Census than the 1999 Census. Furthermore, for the description of house type the questionnaires differ but the information can be made comparable.[2] Variables, relevant for measuring economic performance and available in both censuses include possession of television (hereafter referred to as television), living in a permanent house (henceforth permanent house); using a sanitary toilet (henceforth sanitary toilet); using clean water for living purposes (henceforth clean water); using electricity as the main lighting source (henceforth electricity), being able to read and write (henceforth literacy); possessing a university and above diploma (henceforth university diploma); and possessing any diploma from short-term vocational training (lowest) to PhD (highest). The longer list of durable goods for which we only have information in the 2009 Census are possession of telephone, computer, washing machine, refrigerator, air-conditioner and motorbike.

IV. Comparing VHLSS and non-VHLSS communes

It is clear from the above that the VHLSSs and our data from the 1999 the 2009 censuses are sampled in three different ways. The 33 percent subset of the 1999 Census is a simple random sample while the 15 percent subset of the 2009 Census is stratified at the district level and the VHLSSs are stratified at the province (+urban/rural) level.

Thus, to test if the VHLSSs are representative we post-stratify the two censuses at the province and urban/rural levels to align the census samples with the VHLSS sampling design. Specifically, we introduce $L = 122$ strata in accordance with the 61 provinces existing in 1999 and an urban/rural division of each of the 61 provinces. We also look at the two subsamples consisting of the urban and the rural parts of the provinces. Each of these has 61 strata. As the two samples of the censuses are much larger than the VHLSS samples, and more dispersed geographically, the post-stratification and sub-sampling poses no problems in terms of small or empty strata.

[2] As almost every household has a shelter to live in, we compare a permanent house with semi-permanent and simple houses. In the 1999 Census, respondents were asked and enumerators were requested to make observation to determine whether house was i) permanent; ii) semi-permanent; iii) wood frame of durable use; or iv) simple house (Question 3, Part II). In the 2009 Census, GSO indirectly obtains the house type via 3 questions on construction materials of the main pier (Question 48); construction materials of the roof (Question 49); construction materials of the outer walls (Question 50). Then, a house is classified as permanent if all three components are made with solid materials (CPHCSC 2010).

Let y_{thi} denote the value obtained for one of our outcome measures for household i in stratum h at time t ($t=0$ in 1999 and $t=1$ in 2009). Within each stratum the selection of VHLSS communes is random and the census sample is either random (in 1999) or stratified with known sampling weights (in 2009). Thus, at the stratum level there are three interesting estimators of the population means and proportions. The first estimate is using the complete census sample within each stratum

$$y_{th} = \frac{1}{N_{th}} \sum_{i=1}^{n_{th}} w_{thi} y_{thi}, \quad h = 1, \ldots, L; t = 0,1, \tag{1}$$

where N_{th} is the number of households in stratum h at time t, n_{th} is the sample size (number of observations) in each stratum and w_{thi} is the census specific household weight, which is the inverse of the probability of being selected. For the 1999 Census the weight is the inverse of the sampling fraction for all households in the stratum, $w_{0hi} = N_{0h} / n_{0h}$, such that the estimator is the simple sample average. For the 2009 Census we use the sampling weights provided by GSO. Even though the estimator for the 1999 Census is simply the sample mean, we have chosen to formulate all estimators as Horvitz–Thompson estimators (Horvitz and Thompson, 1952; Cochran, 1977). This underlines the simple common structure of all the estimators we apply.

The two estimates of prime interest are for the VHLSS and the non-VHLSS communes, respectively:

$$y^1_{th} = \frac{1}{N_{th}} \sum_{i=1}^{n_{th}} \mathbf{1}_{VHLSS} \pi_h^{-1} w_{thi} y_{thi}, \quad h = 1, \ldots, L; t = 0,1, \tag{2}$$

$$y^0_{th} = \frac{1}{N_{th}} \sum_{i=1}^{n_{th}} (1 - \mathbf{1}_{VHLSS})(1 - \pi_h)^{-1} w_{thi} y_{thi}, \quad h = 1, \ldots, L; t = 0,1. \tag{3}$$

Here, y^1_{th} is the stratum level estimator of the average or proportion based only on the sampled households living in VHLSS communes while y^0_{th} is the estimator based only on the households living in non-VHLSS communes. In equations (2) and (3) $\mathbf{1}_{VHLSS}$ is the indicator function taking the value 1 if household thi is living in a VHLSS commune and the value 0 if it is living in a non-VHLSS commune and $\pi_h = N^1_{0h} / N_{0h}$ is the probability that the household is living in a VHLSS commune, where N^1_{0h} is the number of households in the VHLSS communes in stratum

h at in 1999. The VHLSS selection weights, π_h, are constant within strata and over time but they vary across strata because of the VHLSS sampling design, described in Section II.

For the 1999 Census, the three estimators reduce to the simple averages of the (sub-) samples, and for both censuses the weighted average of the VHLSS and the non-VHLSS based estimators equals the full sample estimate, using π_h and $(1-\pi_h)$ as weights.

In our test of the representativeness of the VHLSS communes we look at the difference between the two sub-sample estimates

$$d_{th} = y_{th}^1 - y_{th}^0, \quad h = 1,\ldots,L; t = 0,1 \tag{4}$$

As the VHLSS sampling design was simple random sampling at the stratum level, we are comparing two means (or proportions) based on two independent random samples. Hence, we are estimating the stratum level average difference in a randomized controlled trial.

We estimate the overall average difference using stratum weights:

$$d_t = \sum_{h=1}^{L} W_{th} d_{th}, \quad t = 0,1. \tag{5}$$

The stratum weights, W_{th}, are given by the share of households in each stratum relative to the total number of households in Vietnam

$$W_{th} = N_{th} / N_t \quad h = 1,\ldots,L; t = 0,, \tag{6}$$

where N_t is the total number of households at time t (and, thus, the sum of the stratum-level populations). The number of households in each stratum is known from the two censuses, thus the weights are non-random.

Clearly, the average difference between the estimators based on the VHLSS and non-VHLSS subsamples equals the difference of the averages

$$d_t = \sum_{h=1}^{L} W_{th} d_{th} = \sum_{h=1}^{L} W_{th}(y_{th}^1 - y_{th}^0) = \sum_{h=1}^{L} W_{th} y_{th}^1 - \sum_{h=1}^{L} W_{th} y_{th}^0 = y_t^1 - y_t^0, \quad t = 0,1. \tag{7}$$

Therefore, we report all three estimates of the overall averages in our results section in addition to the average difference. We also report averages for the urban and rural sub-samples, respectively. The estimators for the two sub-samples have exactly the same structure as given above, but there are only 61 strata.

Given the two estimates of the differences based on the 1999 Census and the 2009 Census, respectively, it is of central interest to look at the change in the difference, as this is a difference-

in-difference estimate, typically associated with estimation of causal effects in the impact evaluation literature. Hence, we look at the causal effect of being surveyed by the estimate

$$d_1 - d_0 = (y_1^1 - y_1^0) - (y_0^1 - y_0^0). \tag{8}$$

The variances of all averages can be estimated using standard formulae (see *e.g.*, Cochrane, 1977) and variances of the differences are estimated as the sum of the variances of the terms because of the independence, say:

$$v(d_t) = v(y_t^1 - y_t^0) = v(y_t^1) + v(y_t^0) \tag{9}$$

Based on these estimators we test the representativeness of the VHLSS sample and the possible causal effect of being surveyed by classical *t*-ratios. We assume the samples are sufficiently large to allow for the usual normal approximation of the estimators.

V. Results

A. Comparison of Population Estimates in 1999 and 2009

Estimates of the share of households having a certain attribute in 1999, as described in Section III, are given in Table 1. For television, permanent house, sanitary toilet, clean water, and electricity the shares are for the population of households in Vietnam and for the two sub-populations, Rural-Vietnam and Urban-Vietnam, respectively. For literacy and educational diplomas the relevant populations are sub-sets of the Vietnamese population. For literacy, we have defined the relevant population to be all individuals above the age of 9 years. For diploma we have set a minimum age of 18 years and for university diploma we have a minimum age at 22 years. The estimates for the shares of literate and educated people are also given for the two sub-populations of people living in Rural-Vietnam and Urban-Vietnam.

Each block in Table 1 has three estimates of the population shares: the first is based on the complete 33 percent sample of the 1999 Census, the second uses the sub-set of sampled households who resided in communes that were not selected for the VHLSS (non-VHLSS), while the third is based on the remaining households living in the communes that were selected for the VHLSS. Clearly, the best estimate is the one using the full census sample and this estimate is independent of our post-stratification procedure. Hence, the estimates given in the first row of Table 1 are as close as one can get to complete knowledge of the average attributes of the Vietnamese population in 1999.

The most salient result in Table 1 is that the estimates based on households in the VHLSS communes are greater than the estimates based on households in non-VHLSS communes for *all* attributes. Moreover, the differences between the two estimates, given as the 'Difference', are so large, relative to the standard errors, that tests based the normal approximation of the estimators lead us to reject the hypotheses that the differences are zero, at any reasonable level of significance for all attributes. Clearly, the attributes are positively correlated, but the consistency of the sign, size, and significance of the differences strongly indicates that households in the VHLSS communes were better-off than their counterparts in the other communes in 1999.

The division into the two sub-populations, Rural- and Urban-Vietnam, is also telling, both for the situation in Vietnam in 1999 and for the difference between VHLSS and non-VHLSS communes. There are substantial differences in household endowments between Rural and Urban Vietnam. Urban households have much higher likelihoods of having any of the attributes we look at than rural households. For many of the attributes the difference between the urban and rural shares is about 20 percentage points. For sanitary toilets we even record a difference of 50 percentage points as more than 56 percent of the urban households have access to a sanitary toilet while this is so for just about 5 percent of the rural households.

Comparing the estimates based on the VHLSS and non-VHLSS communes we find the former estimates to be significantly larger than the latter in both Rural- and Urban-Vietnam. It is also interesting that for most attributes we find the difference between the estimates to be larger for the rural communes than for the urban communes, despite the much lower shares of households actually possessing the goods, assets and educational attainment in the rural communes. As poverty is predominantly a rural phenomenon in Vietnam this points towards a possibly disturbing bias in the VHLSS based poverty estimates.

Turning to the results for 2009, we have more detailed information about households' endowments as many goods such as motorbikes, telephones, and computers became more common during the 2000s and were thus included in the 2009 Census questionnaire. But, as seen from Table 2, the rapid development in Vietnam during the 2000s is also strikingly visible from the increase in housing conditions and educational attainment also measured in 1999. Most remarkable is the share of households living in a permanent house and the share of households having access to a sanitary toilet. In 1999, 12.8 percent of the households lived in a permanent house; by 2009 the share had increased to 46.7 percent. Even considering the two different

census samples and possible measurement errors, this increase is spectacular. For access to a sanitary toilet the change is of the same order of magnitude from 17.6 percent in 1999 to 53.7 percent in 2009.

Given this development in household attributes over the 10 years the result in Table 2 that all estimates for 2009 based on households living in VHLSS communes are significantly larger than the estimates based on households living in non-VHLSS communes is clearly remarkable given the assumption that the VHLSS commune sample is random at the stratum level. Our finding illustrates that the differences we found for 1999 have persisted to 2009, and the results are underpinned by the addition of the new durable goods.

Furthermore, splitting the samples into Rural- and Urban-Vietnam we find the same pattern of differences in the estimated shares in most cases being larger for Rural-Vietnam than for Urban-Vietnam. A noteworthy exception is the share of households living in a permanent house for which we find that the estimate based on households living in VHLSS communes is actually lower than the estimate based on households living in non-VHLSS communes for Urban-Vietnam.

Even so, the main conclusion from Tables 1 and 2 is that the VHLSS communes appear to be a somewhat extreme sample of communes in the sense that using the households in these communes to form country wide estimates leads to consistent and persistent over-estimation of average household possessions of durable goods, house attributes, and educational attainment, thereby questioning the national representativeness of the information from the 4 rounds of VHLSS in the 2000s.

B. Causal Effects of the VHLSS

The results of Tables 1 and 2 can be combined to give an estimate of the difference over time in the difference between households in VHLSS and non-VHLSS communes. As we do not follow the same households over time we cannot give true difference-in-difference estimates, but our mapping of the communes between 1999 and 2009 creates a pseudo-panel at the commune level. Specifically, if we consider the commune level share of a household attribute, say possession of television, as the parameter of interest, we have estimates of these shares in 1999 and 2009 and it is interesting to look at the difference-in-difference for VHLSS and non-VHLSS communes as this has an interpretation as the causal effect of being selected for the VHLSS survey in 2001.

The difference-in-difference results are reported in Table 3. All results in Table 3 can be computed from Tables 1 and 2 as the estimates are simply the differences of the mean differences given in the tables and the variances of the difference-in-difference results are the sums of the variances of the two single difference estimates given in the tables. The results can be interpreted as the differences between the changes over time in the estimated population shares of the household attributes when using randomly selected households in the VHLSS and the non-VHLSS communes, respectively, to form the population estimates. When the estimated changes differ significantly the VHLSS sample of communes must be considered to be more than an unlikely selection as there appears to be an independent effect of being selected.

In contrast to the results in Tables 1 and 2 we do not have consistent differences between the two samples. For some attributes the changes are larger for households in the VHLSS communes than the other communes while it is reversed for other attributes. In particular, for possession of television and access to electricity we find the non-VHLSS based estimates to have larger changes than the VHLSS-based estimates. Thus relatively more households in non-VHLSS communes have been connected to the power grid and relatively more households have bought televisions. It is not surprising that this goes hand in hand. We also find that literacy rates have increased slightly but nevertheless significantly more in non-VHLSS communes compared to VHLSS communes. These results are interesting because electricity and literacy (in the form of primary education) are two areas of high priority on Vietnam with strong government involvement and we find the shares of both to be very high in 2009; *i.e.*, 96 percent of all households in Vietnam had access to the power grid in 2009 (up from 77.8 percent in 1999) while the literacy rate only increased from 91.1 percent in 1999 to 93.9 percent in 2009. Thus there are many communes for whom the shares were 100 percent in both 1999 and 2009.

However, we find the results for house type, type of toilet and access to clean water in combination with the results for educational attainment in the form of diploma and university diploma to be more important because they to a larger extent reveal costly individual household decisions. And for all of these attributes we find the changes over time to be larger for the households in the VHLSS communes compared to the households in the other communes. We take this to be a strong indication of an independent impact on household well-being of the VHLSS commune selection.

It is interesting to note that the causal effects are driven by the Rural-Vietnam sub-population, while there appears to have been a significant catch-up in the urban sub-population. This indicates that both the levels and the changes in the VHLSS based estimated poverty rates could be severely biased in the sense that poverty levels are probably higher than estimated in 2002, 2004, 2006, and 2008 while the changes in poverty rates from 2002 to 2008 are likely to be lower than estimated.

C. Robustness

Our testing of the representativeness and causal effect of the VHLSS commune selection by comparing population average estimates is guided by the common use of the VHLSS and living standard surveys, and population censuses in general but it is of course not the only way of testing. Thus, to demonstrate that the post-stratification weighting is not driving our results we report the outcomes of another testing strategy in this section.

Specifically, the strata can be considered as giving rise to 122 independent experiments in which each of the 122 populations is randomly divided into two sub-populations; the VHLSS and the non-VHLSS. For each of the experiments we estimate the average difference between household attributes for households living in VHLSS communes and households living in non-VHLSS communes. These experiments generate the attribute share differences, d_{th}, defined in equation (4). But instead of aggregating the differences using stratum weights we can estimate the average difference with the smallest variance. This is a common way of estimating effect sizes in meta-analyses of independent experiments (see *e.g.*, Hedges and Olkin, 1985).

Assuming the true difference is constant across the strata, as it should be a mean zero random variable in all cases if the VHLSS sampling is representative overall, we can use the inverse of the within-stratum variances as weights. Thus, instead of estimating the population averages as in equation (5) we estimate the minimum variance average difference, given by

$$\bar{d}_t = \frac{1}{\sum_{h=1}^{L} \omega_{th}} \sum_{h=1}^{L} \omega_{th} d_{th}, \quad \omega_{th} = \frac{1}{v(d_{th})}, \quad t = 0,1. \tag{10}$$

The variance of this estimator is the inverse of the sum of the weights

$$v(\bar{d}_t) = \frac{1}{\sum_{h=1}^{L} \omega_{th}} = \left(\sum_{h=1}^{L} \frac{1}{v(d_{th})} \right)^{-1}. \tag{11}$$

We also compute stratum-level difference-in-difference estimates, d_{1h} - d_{0h}, and average these exactly as in equation (10) using the inverse of the difference-in-difference variances as weights.

The outcomes of the alternative testing strategy are reported in Tables 4 to 6. The first result to note is that the estimated standard errors of the new average differences are never larger than the standard errors of the differences in the population estimates in the corresponding Tables 1-3. This simply confirms that the new averages are constructed to have smaller variances.

Table 4 shows the weighted average differences for 1999 and the results are comparable to the differences given in Table 1. There is no uniform pattern in the variation between the two estimates, but for most of the attributes the new estimates in Table 4 point towards a smaller average difference than estimated by the population average estimator in Table 1. The largest changes are found for electricity and sanitary toilet for which the minimum variance differences are quite small. However, in terms of statistical significance, the result is unaffected; all differences are extreme in the sense of being very large compared to the standard errors. Using the normal approximation we again find the differences for all attributes to be significant at any reasonable significance level.

The results for 2009, shown in Table 5, follow the same pattern. The minimum variance average differences are generally smaller than the differences in population estimates, but not uniformly so. The only result not following the pattern of equal sign and significance is the difference in the share of households living in a permanent house in Urban-Vietnam, which is the only negative difference in Table 2. In Table 5 the average difference is positive, but statistically insignificant. Thus, overall the new averages clearly support the conclusion that the VHLSS commune sample is extreme, in the sense that households in these communes on average appear to be better off compared to other households.

Finally, the results of the difference-in-difference tests reported in Table 6 support the hypothesis of a causal effect of being in the VHLSS commune sample. The signs of the effects correspond to the estimates in Table 3 in all but one case, clean water in Urban Vietnam, and the changes in the differences are generally statistically significant. Still, most of the new estimates of the differences are smaller than the corresponding estimates in Table 3.

Despite the very large data sets we are using in our tests all average tests reported could potentially be driven by a few large stratum-level differences that dominate the overall averages. To document that this is not the case we present the outcomes of simple sign tests in Tables 7, 8,

and 9. That is, we compute the number of times we observe households in the VHLSS communes have larger stratum level averages than the households in the non-VHLSS communes. These are denoted "positive differences". With 122 strata we expect to have 61 positive (and 61 negative) differences if the VHLSS sample is random in the sense of leading to a median unbiased distribution of the household attributes across the strata. For Rural- and Urban-Vietnam we have 61 strata, thus we expect to find 30.5 positive (and 30.5 negative) differences in each of the sub-populations. In addition to the simple counts of positive differences we also report the probability of observing a more extreme outcome in terms of more positive (or more negative) differences when the data generating mechanism is a binomial distribution with equal probability of positive or negative observations. This is the sign test.

The results in Tables 7 and 8, for the 1999 and 2009 censuses, respectively, document that our main results are not driven by a few outliers. For all attributes the majority of the stratum level differences are positive, and the sign tests reveal that the number of positive differences is too large to support an assumption of equal probabilities of positive and negative differences, *i.e.*, the median difference between average household attributes in VHLSS and non-VHLSS communes is positive. The sample split into urban and rural sub-populations illustrates that the overall result is mainly driven by the rural strata, supporting the outcome of the *t*-tests of the average differences.

For the difference-in-difference tests in Table 9 the number of positive differences and the sign test also support our results in the sense that we find strong indications of negative median effects for electricity and possession of television, driven by the rural stratum level differences, while we find more symmetric distributions of positive and negative changes in the differences for the other household attributes. The opposite signs of the *t*-statistics between the urban and rural areas for permanent houses are supported by the sign tests as we find the majority of rural differences to be positive while the majority of the urban differences are negative, giving rise to an equal number of positive and negative differences at the overall aggregate level. Having a sanitary toilet and possessing a university diploma have similar distributional characteristics.

Overall, we find the *t*-tests and the sign-tests to be in good agreement, indicating that the more powerful parametric test results are not driven by a few outliers.

VI. Conclusion

The Vietnam Household Living Standard Survey is a highly respected survey series providing data to monitor poverty and living standards as well data for a large variety of empirical analyses of interesting developing country topics.

The VHLSS sampling design was planned in the early 2000s. As part of the design, a random draw of about 3,000 communes—roughly one-third of all communes in Vietnam—was only done once, at the initialization of the survey series in 2002. From then onwards the selected communes were unchanged in the succeeding surveys in 2004, 2006, and 2008. Keeping the commune sample unchanged for almost a decade yields administrative benefits as local enumerators can be trained and used more efficiently, yet there is also a risk, well known form other longitudinal surveys, that the surveyed objects are affected by the survey selection.

Exploiting the opportunity of having two Population and Housing Censuses (1999 and 2009) spanning the same decade we examine (i) if the VHLSS commune sample is representative for Vietnam in the sense that households in the VHLSS communes have the same attributes, on average, as households living in the other communes, and (ii) whether the fixed VHLSS commune selection had impacts on the average household attributes during the 2000s.

Our findings indicate that the VHLSS commune selection was somewhat extreme in the sense that average household attributes, *i.e.*, possession of durable goods, quality of housing (including water access and sanitary conditions), and educational attainment, were significantly better in the VHLSS communes compared to non-VHLSS communes. And this was so both in 1999 and in 2009 according to our large samples from the two censuses.

We also find that the VHLSS had an impact on average household welfare. There was a change in the direction of convergence of wellbeing as access to the power grid and possession of television both became relatively more widespread in the non-VHLSS communes over the decade. However, for most attributes that proxy for material well-being and independent household choice, such as housing conditions and educational attainments above primary schooling, the differences between households in VHLSS communes and in non-VHLSS communes increased from 1999 to 2009. In that sense being selected for the VHLSS surveys had an independent positive impact on average household welfare.

We would argue that our findings should cause us to question the validity of the VHLSS-based poverty estimates for Vietnam. If the sampling of communes is extreme, as indicated, the

poverty estimates may be severely downward biased, and the celebrated poverty reduction in Vietnam may be over-estimated. Yet, as the link from the household attributes, given in the censuses, to consumption levels and poverty status is not simple, our results are only indicative. However, if poverty mappings, *i.e.*, small area estimation of poverty rates based on combining household surveys and population censuses, are valid then our results do show that poverty rates in Vietnam are under-estimated.

Given our claim of a causal impact of the survey selection on the wellbeing of the average household in the VHLSS commune sample it is clearly of interest to investigate possible causal mechanisms. We think of three different mechanisms.

First of all, as we are comparing groups over time, not individuals, there is a risk that the result is just a statistical fluke, such as Simpsons Paradox (Simpson, 1951). Another possibility is that the chosen sampling technique, the "every *k*th" rule, is in fact not akin to random sampling when the communes are listed geographically in a grid. Considering the geographical shape of Vietnam one may suspect that household wellbeing is related to spatial placement even within provinces. For example, begin closer to the coast or to the only main road connecting the north and the south, highway A1, may increase the economic opportunities for households. If this is so, then the estimated impact is a result of the VHLSS sampling procedure and it means that the VHLSS estimates should be questioned as they are not based on a random sample. Luckily, the remedy to solve the problem is quite simple as it just requires a new (random) sampling of communes and households in future VHLSS surveys.

The final possible mechanism is that the selected communes actually had increased resource flows compared to the omitted communes. Such increased resource flow could be caused by an awareness of the special attribute of thee communes: the fact that they were measured. We document three cases in which the international research community actually did increase their activities in sub-sets of the selected communes by resurveying households selected for the VHLSS. As donors, during the 2000s, have increasingly demanded statistical proof of aid impact they have also looked for baseline surveys and this is exactly the VHLSS commune attribute that may have caused increased donor activity, and thus resources, in the selected communes. Finally, we cannot reject, nor document, that the Vietnamese authorities have channeled relatively more resources towards the VHLSS communes. However, the results for connection to the power grid and literacy rates do not support the hypothesis

The VHLSS surveys are but one example of LSMS-type household surveys in the developing countries. Thus, based on our results for Vietnam it seems appropriate to reinvestigate the LSMS survey sampling designs in other countries and to test for possible biases in the national poverty estimates for other countries as well. If national poverty rates are under-estimated in more countries this would naturally change our global poverty rates and possibly our perception of our success in reaching the first Millennium Development Goal.

References

Barslund, M., and F. Tarp. 2008. "Formal and Informal Rural Credit in Four Provinces of Vietnam." *The Journal of Development Studies* 44 (4): 485–503.

Baulch, B., Th. K. Ch. Truong, D. Haughton, and J. Haughton. 2004. "Ethnic Minority Development in Vietnam: A Socioeconomic Perspective." In *Economic Growth, Poverty, and Household Welfare in Vietnam*, ed. D. Dollar, P. Glewwe, and N. Agrawal. Washington DC: World Bank.

Bedi, T., A. Coudouel, and K. Simler, ed. 2007. *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*. Washington DC, USA: World Bank.

Benjamin, D., L. Brandt, B. McCaig, and L. H. Nguyen. 2012. *Evaluating the Impact of a Targeted Land Distribution Program: Evidence from Vietnam*.

Chen, S., and M. Ravallion. 2010. "The Developing World Is Poorer Than We Thought, but No Less Successful in the Fight Against Poverty." *The Quarterly Journal of Economics* 125 (4): 1577–1625.

Cochran, W. G. 1977. *Sampling Techniques*. 3rd edition. New York: John Wiley & Sons.

CPHCSC. 2010. *The 2009 Vietnam Population and Housing Census: Major Findings*. Central Population and Housing Census Steering Committee.

Do, Q-T., and T. D. Phung. 2010. "The Importance of Being Wanted." *American Economic Journal: Applied Economics* 2 (4): 236–53. doi:10.1257/app.2.4.236.

Elbers, C., J. O. Lanjouw, and P. Lanjouw. 2003. "Micro–Level Estimation of Poverty and Inequality." *Econometrica* 71 (1) (January): 355–364.

Hansen, H., H. G. Pham, and H. D. Vu. 2007. "A Look at Parameter Constancy and Poverty Predictions in Poverty Mapping Models." In *Market, Policy and Poverty Reduction in Vietnam*, ed. Th. Nguyen and H. Hansen. Hanoi: Vietnam Culture and Information Publishing House.

Hedges, L. V., and I. Olkin. 1985. *Statistical Models for Meta-Analysis*. New York: Academic Press.

Horvitz, D. G., and D. J. Thompson. 1952. "A Generalization of Sampling Without Replacement from a Finite Universe." *Journal of the American Statistical Association* 47 (260): 663–685.

Imai, K.S., R. Gaiha, and W. Kang. 2011. "Poverty, Inequality and Ethnic Minorities in Vietnam." *International Review of Applied Economics* 25 (3): 249–282.

Markussen, T., F. Tarp, and K. Van Den Broeck. 2011. "The Forgotten Property Rights: Evidence on Land Use Rights in Vietnam." *World Development* 39 (5): 839–850.

Mont, D., and C. V. Nguyen. 2011. “Disability and Poverty in Vietnam.” *The World Bank Economic Review* 25 (2): 323–359.

Nakata, H., Y. Sawada, and M. Tanaka. 2009. *Asking Retrospective Questions in Household Surveys: Evidence from Vietnam*. RIETI Discussion Paper Series 10-E.

Nguyen, C. V., Ng. Tr. Tran, and R. Van Der Weide. 2010. “Poverty and Inequality Maps in Rural Vietnam: An Application of Small Area Estimation.” *Asian Economic Journal* 24 (4): 355–390.

Oostendorp, R.H., Q. Tr. Tran, and Th. T. Nguyen. 2009. “The Changing Role of Non-farm Household Enterprises in Vietnam.” *World Development* 37 (3): 632–644.

Phung, D. T., and P. Nguyen. 2007. “Vietnam Household Living Standards Survey (VHLSS) 2002 and 2004: Basic Information.” In *Market, Policy and Poverty Reduction in Vietnam*, ed. Th. Nguyen and H. Hansen, 228–259. Hanoi: Vietnam National Political Publisher.

Ravallion, M., and S. Chen. 2004. “How Have the World’s Poorest Fared Since the Early 1980s?” *World Bank Research Observer* 19 (2): 141–169.

Simpson, E. H. 1951. “The Interpretation of Interaction in Contingency Tables.” *Journal of the Royal Statistical Society. Series B (Methodological)* 13 (2): 238–241.

VASS. 2007. *Poverty Update 2006: Poverty and Poverty Reduction in Vietnam During the Period 1993-2004*. Hanoi: National Political Publishers.

———. 2011. *Poverty Reduction in Vietnam: Achievements and Challenges*. Hanoi: The Gioi Publishing House.

World Bank. 2003. *Vietnam Development Report 2004: Poverty*.

Tables

TABLE 1: ESTIMATED PERCENTAGE SHARE OF THE VIETNAMESE POPULATION HAVING THE ATTRIBUTE. ESTIMATES BASED ON THE 1999 CENSUS

Indicator	Television	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
				All Vietnam				
Overall share	53.73	12.79	17.58	67.92	77.77	91.09	2.45	9.23
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Estimate based on non-VHLSS communes	52.76	12.52	16.87	66.99	76.36	90.73	2.34	9.01
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Estimate based on VHLSS communes	55.36	13.30	18.55	69.83	80.19	91.73	2.58	9.58
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Difference	2.60***	0.79***	1.67***	2.84***	3.82***	1.00***	0.24***	0.57***
	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)
				Rural Vietnam				
Overall share	46.43	8.72	5.19	61.41	72.06	89.87	0.78	5.52
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Estimate based on non-VHLSS communes	45.39	8.47	4.65	60.45	70.41	89.46	0.74	5.38
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Estimate based on VHLSS communes	48.42	9.25	6.04	63.66	75.16	90.68	0.87	5.84
	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.00)	(0.01)
Difference	3.03***	0.78***	1.39***	3.21***	4.76***	1.22***	0.13***	0.46***
	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)
				Urban Vietnam				
Overall share	76.62	25.57	56.46	88.35	95.69	94.93	7.66	20.85
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
Estimate based on non-VHLSS communes	75.90	25.22	55.22	87.50	95.05	94.71	7.37	20.40
	(0.06)	(0.05)	(0.06)	(0.04)	(0.03)	(0.02)	(0.02)	(0.03)
Estimate based on VHLSS communes	77.15	26.02	57.79	89.17	95.94	95.02	7.95	21.31
	(0.04)	(0.04)	(0.04)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)
Difference	1.25***	0.80***	2.56***	1.67***	0.89***	0.30***	0.58***	0.91***
	(0.07)	(0.06)	(0.07)	(0.05)	(0.04)	(0.02)	(0.02)	(0.04)

Notes: Standard errors corrected for sampling design are in parentheses.
Source: Author calculations.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 2: ESTIMATED PERCENTAGE SHARE OF THE VIETNAMESE POPULATION HAVING THE ATTRIBUTE. ESTIMATES BASED ON THE 2009 CENSUS

Indicator	Television	Telephone	Computer	Washing machine	Refrigerator	Air con.	Motorbike	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
							All Vietnam							
Overall share	86.87	45.74	13.53	14.86	31.61	5.90	72.36	46.70	53.72	74.79	96.02	93.86	5.13	14.11
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Estimate based on non-VHLSS communes	86.40	45.27	13.08	14.27	30.84	5.71	71.90	46.47	52.64	73.75	95.59	93.52	4.94	13.75
	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
Estimate based on VHLSS communes	87.84	46.66	14.22	15.84	32.99	6.22	73.05	47.64	55.45	76.68	96.90	94.45	5.43	14.68
	(0.03)	(0.05)	(0.03)	(0.03)	(0.04)	(0.02)	(0.04)	(0.03)	(0.04)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)
Difference	1.44***	1.39***	1.15***	1.56***	2.15***	0.52***	1.15***	1.16***	2.81***	2.93***	1.31***	0.93***	0.49***	0.94***
	(0.04)	(0.06)	(0.05)	(0.05)	(0.06)	(0.03)	(0.05)	(0.05)	(0.05)	(0.04)	(0.02)	(0.01)	(0.02)	(0.03)
							Rural Vietnam							
Overall share	85.19	39.40	6.33	6.55	21.70	1.80	68.26	48.66	41.16	67.57	94.80	92.68	2.23	9.30
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Estimate based on non-VHLSS communes	84.65	38.87	5.90	5.95	20.93	1.61	67.87	48.12	40.06	66.46	94.30	92.32	2.08	9.04
	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.01)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Estimate based on VHLSS communes	86.43	40.58	7.08	7.62	23.28	2.12	69.01	49.98	43.20	69.85	95.97	93.41	2.51	9.86
	(0.04)	(0.06)	(0.03)	(0.03)	(0.05)	(0.02)	(0.05)	(0.04)	(0.05)	(0.04)	(0.02)	(0.01)	(0.01)	(0.02)
Difference	1.77***	1.71***	1.19***	1.67***	2.35***	0.50***	1.13***	1.86***	3.14***	3.40***	1.67***	1.09***	0.43***	0.82***
	(0.05)	(0.07)	(0.04)	(0.04)	(0.06)	(0.02)	(0.06)	(0.05)	(0.06)	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)
							Urban Vietnam							
Overall share	91.56	63.50	33.65	38.09	59.32	17.37	83.82	41.21	88.83	95.00	99.44	97.34	13.64	28.22
	(0.03)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.05)	(0.03)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)
Estimate based on non-VHLSS communes	91.27	63.16	33.14	37.55	58.56	17.15	83.14	41.88	87.81	94.16	99.18	97.05	13.34	27.56
	(0.07)	(0.11)	(0.1)	(0.11)	(0.11)	(0.08)	(0.08)	(0.09)	(0.05)	(0.04)	(0.03)	(0.02)	(0.05)	(0.06)
Estimate based on VHLSS communes	91.79	63.66	34.19	38.81	60.14	17.71	84.35	41.09	89.68	95.77	99.49	97.48	14.02	28.85
	(0.05)	(0.09)	(0.09)	(0.09)	(0.09)	(0.07)	(0.06)	(0.07)	(0.04)	(0.03)	(0.01)	(0.01)	(0.04)	(0.05)
Difference	0.53***	0.49***	1.05***	1.26***	1.58***	0.56***	1.20***	-0.79***	1.87***	1.61***	0.30***	0.43***	0.68***	1.30***
	(0.08)	(0.14)	(0.14)	(0.14)	(0.14)	(0.11)	(0.1)	(0.11)	(0.07)	(0.05)	(0.03)	(0.03)	(0.06)	(0.08)

Notes: Standard errors corrected for sampling design are in parentheses.
Source: Author calculations.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 3: DIFFERENCE-IN-DIFFERENCE: CHANGE FROM 1999 TO 2009 IN DIFFERENCE BETWEEN VHLSS AND NON-VHLSS COMMUNES

Indicator	Television	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
All Vietnam	-1.16***	0.38***	1.13***	0.09*	-2.51***	-0.07***	0.25***	0.37***
	(0.06)	(0.05)	(0.05)	(0.05)	(0.03)	(0.02)	(0.02)	(0.03)
Rural-Vietnam	-1.26***	1.08***	1.75***	0.19***	-3.09***	-0.12***	0.30***	0.35***
	(0.06)	(0.05)	(0.06)	(0.06)	(0.04)	(0.02)	(0.02)	(0.03)
Urban-Vietnam	-0.72***	-1.59***	-0.69***	-0.06	-0.59***	0.12***	0.09	0.39***
	(0.11)	(0.13)	(0.1)	(0.07)	(0.05)	(0.03)	(0.07)	(0.08)

Notes: Standard errors corrected for sampling design are in parentheses.
Source: Author calculations.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 4: WEIGHTED AVERAGE STRATUM-LEVEL DIFFERENCE BETWEEN VHLSS AND NON-VHLSS COMMUNES, 1999

Indicator	Television	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
All Vietnam	2.35***	0.59***	0.65***	1.86***	0.53***	0.63***	0.09***	0.30***
	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.)	(0.01)
Rural-Vietnam	3.09***	0.58***	0.53***	3.15***	1.17***	0.89***	0.07***	0.27***
	(0.04)	(0.02)	(0.01)	(0.03)	(0.02)	(0.01)	(0.)	(0.01)
Urban-Vietnam	0.93***	0.79***	2.79***	0.74***	0.12***	0.16***	0.43***	0.55***
	(0.06)	(0.06)	(0.06)	(0.03)	(0.01)	(0.02)	(0.02)	(0.03)

Notes: Weighted averages of stratum differences using the inverse of the stratum difference variance as weight. Standard errors in parentheses are the square roots of the harmonic means of the stratum difference variances.
Source: Author calculations.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 5: WEIGHTED AVERAGE STRATUM-LEVEL DIFFERENCE BETWEEN VHLSS AND NON-VHLSS COMMUNES, 2009

Indicator	Television	Telephone	Computer	Washing machine	Refrigerator	Air con.	Motorbike	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
All Vietnam	1.40***	1.56***	0.66***	0.69***	2.07***	0.09***	1.24***	0.86***	1.93***	0.34***	0.25***	0.62***	0.24***	0.66***
	(0.03)	(0.05)	(0.02)	(0.02)	(0.04)	(0.01)	(0.05)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Rural-Vietnam	1.86***	1.89***	0.62***	0.63***	2.07***	0.08***	1.30***	1.02***	3.46***	2.76***	0.39***	0.91***	0.23***	0.58***
	(0.04)	(0.06)	(0.03)	(0.02)	(0.05)	(0.01)	(0.06)	(0.04)	(0.05)	(0.04)	(0.01)	(0.02)	(0.01)	(0.02)
Urban-Vietnam	0.59***	0.53***	1.34***	2.17***	2.06***	0.33***	1.11***	0.12	0.87***	0.09***	0.08***	0.26***	0.52***	1.41***
	(0.06)	(0.11)	(0.1)	(0.11)	(0.11)	(0.05)	(0.08)	(0.08)	(0.04)	(0.01)	(0.01)	(0.02)	(0.05)	(0.06)

Notes: Weighted averages of stratum differences using the inverse of the stratum difference variance as weight. Standard errors in parentheses are the square roots of the harmonic means of the stratum difference variances.
Source: Author calculations.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 6: WEIGHTED AVERAGE STRATUM-LEVEL DIFFERENCE OF CHANGE FROM 1999 TO 2009 IN DIFFERENCE BETWEEN VHLSS AND NON-VHLSS COMMUNES

Indicator	Television	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
All Vietnam	-1.13***	0.33***	0.72***	0.08**	-0.49***	-0.14***	0.12***	0.19***
	(0.05)	(0.04)	(0.05)	(0.03)	(0.01)	(0.02)	(0.01)	(0.02)
Rural-Vietnam	-1.31***	0.59***	1.75***	-0.06	-0.85***	-0.26***	0.12***	0.15***
	(0.06)	(0.04)	(0.06)	(0.05)	(0.02)	(0.02)	(0.01)	(0.02)
Urban-Vietnam	-0.75***	-1.14***	-1.23***	0.16***	-0.13***	0.03	0.07	0.53***
	(0.09)	(0.10)	(0.08)	(0.04)	(0.02)	(0.03)	(0.06)	(0.07)

Notes: Weighted averages of stratum differences using the inverse of the stratum difference-in-difference variance as weight. Standard errors in parentheses are the square roots of the harmonic means of the stratum difference-in-difference variances.
Source: Author calculations.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 7: NUMBER OF POSITIVE STRATUM LEVEL DIFFERENCES BETWEEN VHLSS AND NON-VHLSS COMMUNES, 1999 CENSUS

Indicator	Television	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
All Vietnam	90	84	85	87	86	83	79	74
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Rural-Vietnam	52	49	47	49	52	48	47	42
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Urban-Vietnam	38	35	38	38	34	35	32	32
	0.07	0.31	0.07	0.07	0.44	0.31	0.80	0.80

Notes: First, (third and fifth) row gives the number of positive stratum level differences between VHLSS and non-VHLSS communes out of the 122 (61) strata. Second, fourth, and sixth row gives the probability of observing more positive (or negative) differences when the data is generated by a binomial distribution with probability of success equal to 0.5.
Source: Author calculations.

TABLE 8: NUMBER OF POSITIVE STRATUM LEVEL DIFFERENCES BETWEEN VHLSS AND NON-VHLSS COMMUNES, 2009 CENSUS

Indicator	Television	Telephone	Computer	Washing machine	Refrigerator	Air con.	Motorbike	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
All Vietnam	78	71	77	80	81	70	75	75	78	86	85	79	76	75
	0.00	0.08	0.00	0.00	0.00	0.12	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01
Rural-Vietnam	47	43	41	44	46	39	42	45	47	51	47	46	44	41
	0.00	0.00	0.01	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Urban-Vietnam	31	28	36	36	35	31	33	30	31	35	38	33	32	34
	1.00	0.61	0.20	0.20	0.31	1.00	0.61	1.00	1.00	0.31	0.07	0.61	0.80	0.40

Notes: First, (third and fifth) row gives the number of positive stratum level differences between VHLSS and non-VHLSS communes out of the 122 (61) strata. Second, fourth, and sixth row gives the probability of observing more positive (or negative) differences when the data is generated by a binomial distribution with probability of success equal to 0.5.
Source: Author calculations.

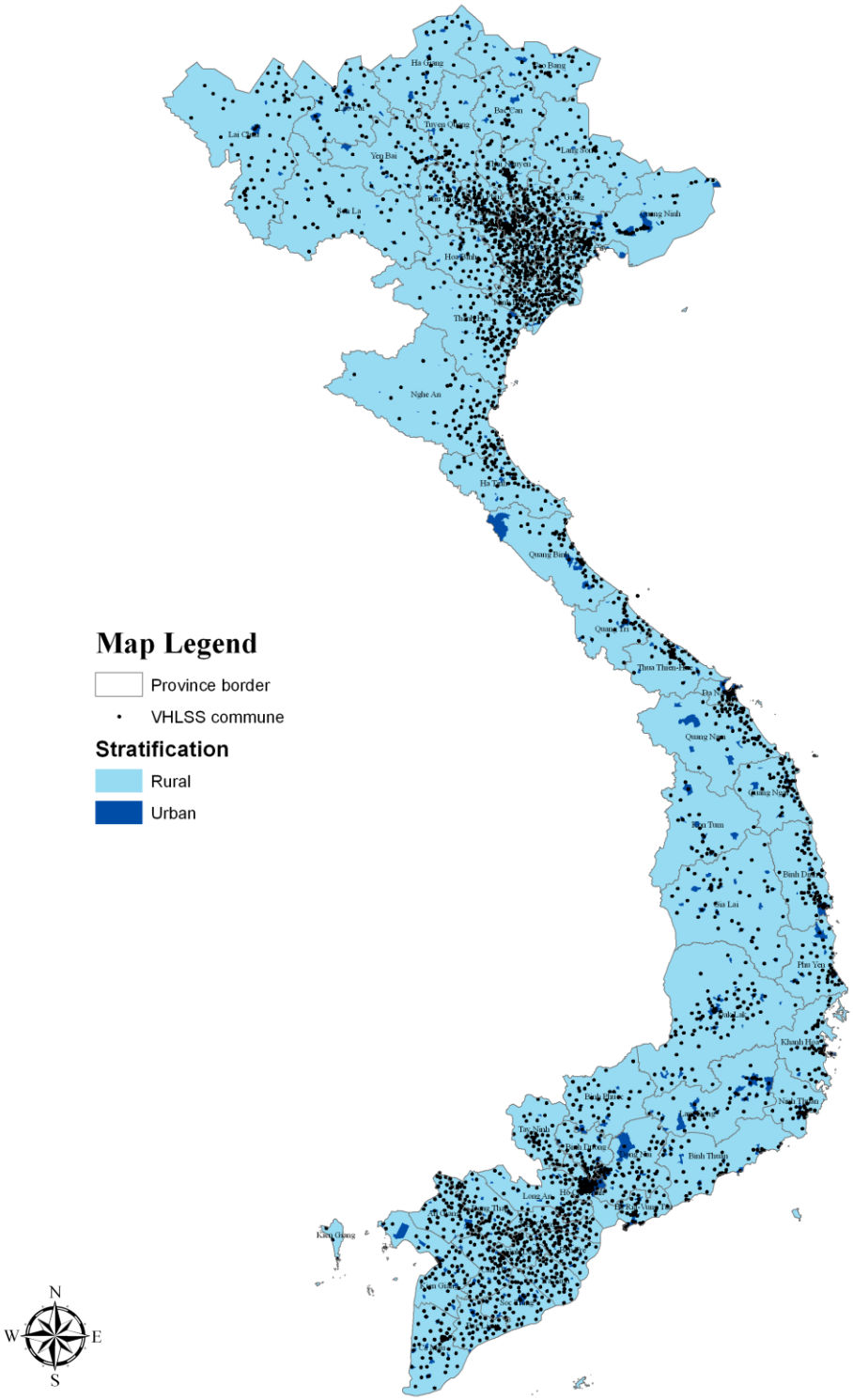
TABLE 9: NUMBER OF POSITIVE STRATUM LEVEL CHANGES FROM 1999 TO 2009 IN THE DIFFERENCES BETWEEN VHLSS AND NON-VHLSS COMMUNES. 1999 AND 2009 CENSUS

Indicator	Television	Permanent house	Sanitary toilet	Clean water	Electricity	Literacy	University diploma	Diploma
All Vietnam	42	61	68	66	44	56	70	68
	0.00	1.00	0.24	0.42	0.00	0.42	0.12	0.24
Rural-Vietnam	19	37	40	31	18	22	41	35
	0.00	0.12	0.02	1.00	0.00	0.04	0.01	0.31
Urban-Vietnam	23	24	28	35	26	34	29	33
	0.07	0.12	0.61	0.31	0.31	0.44	0.80	0.61

Notes: First, (third and fifth) row gives the number of positive stratum level changes from 1999 to 2009 in the differences between VHLSS and non-VHLSS communes out of the 122 (61) strata. Second, fourth, and sixth row gives the probability of observing more positive (or negative) differences when the data is generated by a binomial distribution with probability of success equal to 0.5.
Source: Author calculations.

Figures

Figure 1: The geographical placement of the 3000 VHLSS communes



Source: Authors' construction