The Spatial Distribution of Poverty in Vietnam and the Potential for Targeting

Nicholas Minot and Bob Baulch

April 2002

Contact information: Nicholas Minot is a Research Fellow at the International Food Policy Research Institute (IFPRI), 2033 K Street N.W., Washington, D.C. 20006 U.S.A., email: n.minot@cgiar.org. Bob Baulch is a Fellow at the Institute of Development Studies, University of Sussex and formerly Quantitative Poverty Specialist at the World Bank, Vietnam, email: b.baulch@ids.ac.uk. Senior authorship is not assigned.

Acknowledgements: We thank Phan Xuan Cam and Nguyen Van Minh for their help understanding the Vietnam Census data and Peter Lanjouw for helpful methodological discussions. Paul Glewwe and participants at workshops in Hanoi produced valuable comments on earlier versions this paper. The financial assistance of the DFID Poverty Analysis and Policy support Trust Fund and World Bank Development Economics research Group is acknowledged. The usual disclaimers apply.

Table Of Contents

List Of Tables

List Of Figures

Abstract

This paper combines household survey and census data to construct a provincial poverty map of Vietnam and evaluate the accuracy of geographically targeted anti-poverty programs. First, the paper estimates per capita expenditure as a function of selected household and geographic characteristics using the 1998 Vietnam Living Standards Survey. Next, these results are combined with data on the same household characteristics from the 1999 Census to estimate the incidence of poverty in each province. The results indicate that rural poverty is concentrated in ten provinces in the Northern Uplands, two provinces of the central Highlands, and two provinces in the Central Coast. Finally, Receiver Operating Characteristics curves are used to evaluate the effectiveness of geographic targeting. The results show that the existing poor communes system excludes large numbers of poor people, but there is potential to sharpen poverty targeting using a small number of easy-to-measure household characteristics.

1. Introduction

1.1 Background

In most countries, poverty is spatially concentrated. Extreme poverty in inaccessible areas with unfavorable terrain often coexists with relative affluence in more favorable locations close to major cities and markets. Information on the spatial distribution of poverty is of interest to policymakers and researchers for a number of reasons. First, it can be used to quantify suspected regional disparities in living standards and identify which areas are falling behind in the process of economic development. Second, it facilitates the targeting of programs whose purpose is, at least in part, to alleviate poverty such as education, health, credit, and food aid. Third, it may shed light on the geographic factors associated with poverty, such as mountainous terrain or distance from major cities.

Traditionally, information on poverty has come from household income and expenditure surveys. These surveys generally have sample sizes of 2000 to 8000 households, which only allow estimates of poverty for 3 to 12 regions within a country. Previous research has, however, shown that geographic targeting is most effective when the geographic units are quite small, such as a village or district (Baker and Grosh, 1994; Bigman and Fofack, 2000). The only household information usually available at this level of disaggregation is census data, but census questionnaires are generally limited to household characteristics and rarely include questions on income or expenditure.

In recent years, new techniques have been developed that combine household and census data to estimate poverty for more disaggregated geographic units. Although various approaches have been used, they all involve two steps. First, household survey data is used to estimate poverty or expenditure as a function of household characteristics such as household composition, education, occupation, housing characteristics, and asset ownership. Second, census data on those same household characteristics are inserted into the equation to generate estimates of poverty for small geographic areas.

For example, Minot (1998 and 2000) used the 1992-93 Vietnam Living Standards Survey and a probit model to estimate the likelihood of poverty for rural households as a function of a series of household and farm characteristics. District-level means of these same characteristics were then obtained from the 1994 Agricultural Census and inserted into this equation, generating estimates of rural poverty for each of the 543 districts in the country.

Hentschel *et al.* (2000) developed a similar method using survey and census data from Ecuador. Using log-linear regression models and household-level data from a census, they demonstrate that their estimator generates unbiased estimates of the poverty headcount and show how to calculate the standard error of the poverty headcount.[1] This approach has been applied in a number of other countries including Panama and South Africa (see World Bank, 2000; Statistics South Africa and the World Bank, 2000).

The earlier Vietnam study has several limitations. First, since it relied on the Agricultural Census, it generated poverty estimates only for the rural areas. Second, the use of a probit regression and district-level means, although intuitively plausible, does not necessarily generate consistent estimates of district-level poverty[2]. Third, in the absence of household-level census data, it was not possible to estimate the standard errors of the estimates to evaluate their accuracy.

1.2 Objectives

Accordingly, this paper has three objectives. First, it explores the household factors associated with poverty in Vietnam using the 1998 Vietnam Living Standards Survey (VLSS). In this task,

---

1 The poverty headcount is defined as the proportion of the population with per capita expenditures below the poverty line.

2 Minot and Baulch (2002) show that using aggregated census data underestimates the incidence of poverty when it is below 50 percent and overestimates it when it is above 50 percent. The absolute size of the error, however, can be as low as 2-3 percentage points in some circumstances.

it builds on an earlier report describing the characteristics of poor households in Vietnam (Poverty Working Group, 1999).

Second, it examines the spatial distribution of poverty in Vietnam using the 1998 VLSS and a 3 percent sample of the 1999 Population and Housing Census. This analysis represents an improvement on the earlier Vietnam study in several respects: a) the data are more recent, an important consideration in a rapidly growing country such as Vietnam, b) the analysis covers both urban and rural areas, providing a broader view of poverty in Vietnam, and c) we calculate the standard error of the poverty headcount. The standard errors are based on the methods suggested by Hentschel *et al.* (2000), with extensions to incorporate the sampling error associated with the fact that we are using a 3% sample of the Population Census rather than the full Census.

Third, this study examines the efficacy of Vietnam's existing geographically targeted anti-poverty programs and investigates the potential for improving the targeting of the poor by using the type of additional household level variables that could be collected in a "quick-and-dirty" enumeration of households.

1.3 Organization of paper

Section 2 describes the data and methods used to generate poverty maps for Vietnam from household survey data and census data. Section 3 describes the results of the regression analysis. Although these are an input in the poverty mapping procedure, they also yield insights on the factors associated with poverty and how they vary between urban and rural areas. Section 4 presents the provincial estimates of urban and rural poverty in Vietnam, along with the standard errors of these estimates. Section 5 examines the efficacy of Vietnam's poor and disadvantaged communes program and investigates whether use of additional household variables might improve poverty targeting. Finally, Section 6 summarizes the results, discusses some of their policy implications, and suggests areas for future research.

2. Data and Methods

2.1 Data

This study makes use of two data sets: the 1998 Vietnam Living Standards Survey (VLSS) and the 1999 Population and Housing Census. The VLSS was implemented by the General Statistics Office (GSO) of Vietnam with funding from the Swedish International Development Agency and the United Nations Development Program and with technical assistance from the World Bank. The sample included 6000 households (4270 in rural areas and 1730 in urban areas), in Vietnam, selected using a stratified random sample.

The 1999 Census was carried out by the GSO and refers to the situation as of April 1, 1999. It was conducted with the financial and technical support of the United Nations Family Planning Association and the United Nations Development Program. As the full results of the Census have not yet been released, this analysis is based on a 3 percent sample of the Census. The 3 percent sample was selected by GSO using a stratified random sample of 5287 enumeration units and 534,139 households. The 3 percent sample of the Census was designed to be representative at the provincial level.

There are a number of variables which are common to both the VLSS and the Census, and which allow household level expenditures to be predicted and disaggregated poverty estimates produced. Table 1 summarizes the 17 variables that were selected for inclusion in our poverty mapping exercise.

Table 1. Household characteristics common to the Census and the VLSS

Variable name(s)	Description of Variable	Question number 1999 Census	1998 VLSS
hhsize	Household size (number of people)	Pt I,Q4	S1A
pelderly	Proportion of elderly people (aged over 60) in household	Pt I, Q4	S1A,Q2
pchild	Proportion of children (aged under 15) in household	Pt I, Q4	S1A,Q6
pfemale	Proportion of females in household	Pt I, Q3	S1A,Q6
Iedchd_1 to 6	Highest level of education completed by head (less than primary school primary school, lower secondary school, upper secondary school technical or vocation training, college diploma or university degree)	Pt I,Q11-13	S2A
Iedcsp_0	Dummy for no spouse	Pt I, Q2	S1B,Q3
Iedcsp1 to 6	Highest level of education completed by spouse (less thqn primary, primary school, lower secondary school, upper secondary school technical or vocation training, college diploma or university degree)	Pt I,Q11-13	S2A
ethnic	Dummy for ethnic minority head (not Kinh or Chinese)	Pt I, Q4	S0A
Ioccup_1 to 7	Occupation of head over last 12 months (political leader or manager, professional or technical worker, clerk or service worker, agriculture non-farm enterprises, unskilled worker, not-working)	Pt I, Q16	S4D
Ihouse_1 to 3	Type of house (permanent; semi-permanent or wooden frame, "simple"	Pt III, Q3	S6A,Q1
htypla1 to 2	House type interacted with living area (m2)	Pt III, 4	S6C,Q1a
electric	Household with electricity	Pt III, Q7	S6B,Q33
Iwater_1 to 3	Main source of drinking water (private or public tap, rainwater and wel rivers and lakes)	Pt III, 8	S6B,Q25
Itoilet_1 to 3	Type of toilet (flush, latrine/other, none)	Pt III, Q9	S6B,Q31
tv	Dummy for TV ownership	Pt III, Q10	S12C
radio	Dummy for radio ownership	Pt III, Q11	S12C
reg7_1 to 7	Regional dummies (7 regions)	page 1	S0A

Source: Questionnaires for 1998 VLSS and 1999 Population and Housing Census

To estimate the poverty headcount, we predict expenditures using these common variables and then apply the food and overall poverty lines developed by the GSO and the World Bank for use with the VLSS surveys (Poverty Working Group, 1999). The lower of these two lines, the food poverty line, corresponds to the expenditure (including the value of home production and adjusted regional and seasonal price differences) required to purchase 2100 kilocalories per person per day. The upper overall poverty line also incorporates a modest allowance for non-food expenditures.[3]

The Ministry of Labor, Invalids, and Social Assistance (MOLISA) estimates provincial poverty rates based on a system of administrative reporting that uses different welfare indicators (rice equivalent income), different poverty lines, and a different unit of analysis (households). Nonetheless, the results are fairly similar to those obtained in this study.

2.2 Estimating poverty with a household survey

As mentioned above, the first step in implementing this approach is to estimate poverty or household welfare as a function of household characteristics. In this study, we use per capita consumption expenditure as the measure of household welfare. The explanatory variables must be useful in "predicting" household welfare and they must exist in both the household survey and the census. Economic theory provides no guidance on the functional form, but often a log-linear function is used:

$$\ln(y_i) = X'_i\beta + \varepsilon_i \qquad \textbf{(1)}$$

where y_i is the per capita consumption expenditure of household i, X'_i is a kx1 vector of household characteristics of household i, β is a kx1 vector of coefficients, and ε_i is a random

[3] In 1998, the food poverty line was VND 1286,833 and the overall poverty line was VND1,789,871 per person per year. See Annex 2 of Poverty Working Group (1999) for further details concerning the estimation of these poverty lines.

disturbance term distributed as N(0,σ). Because our main interest is predicting the value of ln(y) rather than assessing the impact of each explanatory variable, we are not concerned about the possible endogeneity of some of the explanatory variables.

Hentschel *et al.* (2000) show that the probability that household i with characteristics X_i is poor can be expressed as:

$$E[P_i \mid X_i, \beta, \sigma^2] = \Phi\left[\frac{\ln z - X_i\beta}{\sigma}\right] \quad \textbf{(2)}$$

where P_i is a variable taking a value of 1 if the household is poor and 0 otherwise, z is the poverty line expressed in terms of consumption expenditure per capita, and Φ is the cumulative standard normal function.

2.3 Applying regression results to the census data

In the second step, the estimated regression coefficients from the first step are combined with census data on the same household characteristics to predict the probability that each household in the Census is poor. This is accomplished by inserting the household characteristics for household i from the census, X_i^C , into equation 2:

$$E[P_i \mid X_i, \beta, \sigma^2] = \Phi\left[\frac{\ln z - Xi^C\beta}{\sigma}\right] \quad \textbf{(3)}$$

For a given area (such as a district or province), Hentschel et al (2000) show that the proportion of the population living in households that are below the poverty line is estimated as the mean of the probabilities that individual households are poor:

$$E[P \mid X^C, \beta, \sigma^2] = \sum_{i=1}^{N} \frac{m_i}{M} \Phi\left[\frac{\ln z - X_i^C\beta}{\sigma}\right] \quad \textbf{(4)}$$

where m_i is the size of household i, M is the total population of the area in question, N is the number of households, and X is an N x k matrix of household characteristics. The advantage of using the Census data, of course, is that the large number of households allows estimation of poverty headcounts for geographic units much smaller than would be possible with the VLSS data.

Provided that a) the error term is homoskedastic, b) there is no spatial auto-correlation, and c) the full Census data are used, the variance of the estimated poverty headcount can be calculated as follows:

$$\mathrm{var}(P^*) = \left(\frac{\partial P^*}{\partial \hat{\beta}}\right)' \mathrm{var}(\hat{\beta}) \frac{\partial P^*}{\partial \hat{\beta}} + \left(\frac{\partial P^*}{\partial \hat{\sigma}^2}\right)^2 \frac{2\hat{\sigma}^4}{n-k-1} + \sum_{i=1}^{N} \frac{m_i^2 P_i^* (1-P_i^*)}{M^2} \qquad \textbf{(5)}$$

where n is the sample size in the regression model. Thus, n, k, and σ^2 are from the regression analysis, while m_i, M, and N are obtained from the census data. The partial derivatives of P* with respect to the estimated parameters can be calculated as follows:

$$\frac{\partial P^*}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} \frac{m_i}{M} \left(\frac{-x_{ij}}{\hat{\sigma}}\right) \phi\left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}}\right) \qquad \textbf{(6)}$$

$$\frac{\partial P^*}{\partial \hat{\sigma}^2} = -\frac{1}{2} \sum_{i=1}^{N} \frac{m_i}{M} \left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}^3}\right) \phi\left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}}\right) \qquad \textbf{(7)}$$

The first two terms in equation 5 represent the “model error”, which comes from the fact that there is some uncertainty regarding the true value of β and σ in the regression analysis. This uncertainty is measured by the estimated covariance matrix of β and the estimated variance of σ^2, as well the effect of this variation on P*. The third term in equation 5 measures the “idiosyncratic error” which is related to the fact that, even if β and σ are measured exactly, household-specific factors will cause the actual expenditure to differ from predicted expenditure. These equations are described in more detail in Hentschel *et al.* (2000) and Elbers *et al.* (2001).

As noted above, equation 5 is valid only if the full Census data are available for the second stage of the mapping procedure. When we are using a sample survey or a sample of the Census data in the second stage, this expression must be modified as follows:

$$\operatorname{var}(P^*) = \left(\frac{\partial P^*}{\partial \hat{b}}\right)' \operatorname{var}(\hat{b}) \frac{\partial P^*}{\partial \hat{b}} + \left(\frac{\partial P^*}{\partial \hat{s}^2}\right)^2 \frac{2\hat{s}^4}{n-k-1} + \sum_{i=1}^{N} \frac{m_i^2 P_i^*(1-P_i^*)}{M^2} + V_s \qquad \textbf{(8)}$$

where V_s represents the variance associated with the sampling error in the Census, taking into account the design of the sample. In this study, we rely on the software package Stata to calculate the variance associated with the sampling error, taking into account the design of the survey[4].

In order to compare poverty headcounts in different regions or provinces, it is convenient to calculate the variance of the difference between two estimates of poverty. Hentschel et al (2000, footnote 17) provide an expression for the case when full Census data are used. Here we extend the expression to include the variance associated with sampling error:

$$\begin{aligned}\operatorname{var}(P_1 - P_2) = &\left(\frac{\partial P_1 - P_2}{\partial \hat{b}}\right)' \operatorname{var}(\hat{b}) \left(\frac{\partial P_1 - P_2}{\partial \hat{b}}\right) + \left(\frac{\partial P_1 - P_2}{\partial \hat{s}^2}\right)^2 \frac{2\hat{s}^4}{n-k-1} \\ &+ V_i(P_1) + V_i(P_2) + V_s(P_1) + V_s(P_2) - 2\operatorname{cov}_s(P_1, P_2)\end{aligned} \qquad \textbf{(9)}$$

where $V_i(P_r)$ is the idiosyncratic variance of the poverty estimate for region r (the third term in equation 5), $V_s(P_r)$ is the sampling variance of the poverty estimate for region r, and $\operatorname{cov}_s(P_1,P_2)$ is the covariance in the poverty estimates for regions 1 and 2 associated with sampling error.

4 This is accomplished with the "svymean" command. Stata calculates a linear approximation (a first-order Taylor expansion) of the sampling error variance based on information on the strata, the primary sampling unit, and the weighting factors. See Stata Corporation, 2001b for more information.

Two qualifications need to be made regarding the implementation of this method in the case of Vietnam. Researchers at the World Bank have recently been addressing the issue of spatial autocorrelation in the first-stage regressions (equation 1). Analytical solutions for the variance of the headcount are not possible in this case, and it becomes necessary to use complex simulation methods to calculate the estimators and their standard errors (Elbers *et al*, 2001). Although preliminary analysis indicates the presence of some spatial autocorrelation, we were not able to eliminate it by including community-level variables in the regression analysis. This suggests that there may be some inefficiency in the results of the first-stage regression analysis, though the magnitude of these effects is difficult to assess.

In addition, the estimate of the variance associated with sampling error produced by Stata is only an approximation. Exploratory analysis reveals that the sampling error is relatively small compared to the model error, suggesting that this approximation does not influence the results substantively.

3. Factors Associated with Poverty in Vietnam

As described in Section 2.2, the first step in constructing a poverty map is to estimate econometrically per capita consumption expenditure as a function of variables that are common to the Census and the VLSS. These household characteristics include household size and composition, ethnicity, education of the head of household and his/her spouse, occupation of the head of household, housing size and type, access to basic services, and ownership of selected consumer durables. Table 1 lists the variables and Annex 1 provides descriptive statistics for each of them.

It is reasonable to expect that the factors which “predict” expenditure in rural areas may be different than those predicting expenditure in urban areas. Indeed, a Chow test strongly rejects the hypothesis that the coefficients for the urban sub-sample are the same as those for the rural sub-sample ($F=6.16$, $p< .001$). This implies that we should carry out separate analyses on rural and urban samples.

The next level of disaggregation is the stratum used in the VLSS sample. The VLSS was designed to be representative for each of ten strata, comprising three urban strata and seven rural strata. For this analysis, it was necessary to collapse the three urban strata (Hanoi/Ho Chi Minh City, other cities, and towns) into two (Hanoi/Ho Chi Minh City and other urban areas) because the Census data do not allow us to distinguish between "other cities" and towns. Within urban areas, a Chow test suggests that Hanoi and Ho Chi Minh City differ significantly from other urban areas ($F=2.20$, $p<.001$). In addition, the seven rural regions differ significantly from each other ($F=12.61$, $p<.001$). In other ways, however, the stratum-level regressions are not very satisfactory. Because of the small sample size in each stratum (ranging from 368 to 1111 households), many of the coefficients are not statistically significant at conventional levels or have counter-intuitive signs. Furthermore, the goodness-of-fit of most of the stratum regressions is below 0.5, compared to 0.54 and 0.55 for the rural and urban regressions. One result of this is that the standard errors of the poverty estimates from the stratum-level regressions are higher than those obtained from the urban-rural regressions (see Section 4.1).

In this paper, we will present the results of both the urban-rural regressions (see Tables 2 and 3) and the stratum-level regressions (see Annexes 2 and 3), as well as the poverty estimates derived from each (Tables 4-6 and Annex 4). However, we will give greater prominence to the results from the urban-rural regression analysis. As will be shown later, the two methods yield similar poverty headcounts and rankings, particularly for the poorest provinces. In the six sub-sections that follow, we summarize the results of the regression analysis to "predict" per capita expenditures.

3.1 Household size and composition

Large households are strongly associated with lower per capita expenditure in both urban and rural areas, as shown in Table 2. The negative sign of the coefficient on household size implies that, other factors being equal, each additional household member is associated with a 7-8 percent reduction in per capita expenditure[5]. The stratum-level regressions show similar results (see Annex 2).

In rural areas, a household with a large number of elderly members, of children, and of females is likely to have low per capita expenditure. In urban areas, however, only the number of children is statistically significant (see Table 2). Household composition appears to matter less in urban areas than rural ones. It may be that the number of children, women, and elderly people have less effect on household welfare in urban areas because income-earning capacity in the cities and town is less dependent on physical strength.

Ethnicity[6] is a predictor of per capita expenditure, but a surprisingly weak one. In rural areas, the coefficient on ethnicity was significant only at the 10 percent level while in urban areas, it was not statistically significant (see Table 2). The urban coefficient is not surprising given the very small sample of ethnic minority households in urban areas (just 19 households). The weakly significant, although appropriately signed, coefficient for rural areas is more surprising given the strong correlation between poverty and ethnicity in Vietnam. Other research (Van de Walle and Gundewardana, 2000, Baulch *et al.*, forthcoming) suggests that ethnic minorities have both lower levels of endowments and lower returns to those endowments. Our results are consistent with these findings,

---

[5] A coefficient of –0.772 implies that a one-unit increase in the explanatory variable is associated with 7.4 percent reduction in per capita expenditure, since exp(-0.772)=0.926=1-7.4%. We must be careful before inferring that larger households are worse off than smaller ones, however, for two reasons. First, there may be economies of scale in household size, so that larger households do not “need” the same per capita expenditures as smaller households to reach an equivalent level of welfare. Second, our measure of welfare does not take into account household composition, so if larger households have more children than smaller households they might still have equivalent levels of expenditure *per adult* equivalent.

[6] In common with other studies of ethnic minority issues using the VLSS, we group *Hoa* (Chinese) households along with the *Kinh* (ethnic Vietnamese).

Table 2. Determinants of per capita expenditure for rural and urban areas

Rural model				Urban model			
N	4269				1730		
R-squared	0.536				0.550		
Variable	Coefficient	t		Variable	Coefficient	t	
hhsize	-0.0772	-19.5	***	hhsize	-0.0785	-8.1	***
pelderly	-0.0831	-2.4	**	pelderly	-0.1026	-1.6	
pchild	-0.3353	-9.4	***	pchild	-0.2368	-3.6	***
pfemale	-0.1177	-3.5	***	pfemale	0.0386	0.5	
ethnic	-0.0765	-1.9	*	ethnic	0.0142	0.2	
Iedchd_2	0.0585	3.4	***	Iedchd_2	0.0616	1.7	
Iedchd_3	0.0883	4.5	***	Iedchd_3	0.0338	1.3	
Iedchd_4	0.0884	3.3	***	Iedchd_4	0.1368	3.2	***
Iedchd_5	0.1355	4.2	***	Iedchd_5	0.1603	3.5	***
Iedchd_6	0.2552	4.9	***	Iedchd_6	0.1843	3.7	***
Iedcsp_0	0.0173	1.0		Iedcsp_0	0.0344	0.8	
Iedcsp_2	0.0049	0.3		Iedcsp_2	0.0642	1.9	*
Iedcsp_3	0.0132	0.6		Iedcsp_3	0.0987	2.6	**
Iedcsp_4	0.0107	0.3		Iedcsp_4	0.1912	2.7	**
Iedcsp_5	0.0921	2.3	**	Iedcsp_5	0.1285	3.2	***
Iedcsp_6	0.1571	2.7	***	Iedcsp_6	0.1752	3.1	***
Ioccup_1	0.1414	3.5	***	Ioccup_1	0.2312	3.0	***
Ioccup_2	0.1350	3.3	***	Ioccup_2	0.0576	1.2	
Ioccup_3	0.1362	3.4	***	Ioccup_3	0.0357	0.9	
Ioccup_4	-0.0163	-0.6		Ioccup_4	-0.0093	-0.2	
Ioccup_5	0.0701	1.9	*	Ioccup_5	0.0071	0.2	
Ioccup_6	-0.0586	-1.7	*	Ioccup_6	-0.1599	-2.9	***
Ihouse_1	-0.9228	-4.3	***	Ihouse_1	-0.5194	-3.4	***
Ihouse_2	-0.3120	-3.6	***	Ihouse_2	-0.4001	-3.8	***
htypla1	0.2958	5.7	***	htypla1	0.2001	5.4	***
htypla2	0.1180	5.2	***	htypla2	0.1403	4.6	***
electric	0.0765	2.7	***	electric	-0.0026	0.0	
Inwate_1	0.0828	1.4		Inwate_1	0.2289	5.3	***
Inwate_2	0.1157	4.4	***	Inwate_2	0.0340	0.6	
Itoile_1	0.2700	5.5	***	Itoile_1	0.1311	2.2	**
Itoile_2	0.0556	2.6	**	Itoile_2	0.0049	0.1	
tv	0.2124	15.1	***	tv	0.2167	5.5	***
radio	0.1009	7.0	***	radio	0.1599	6.2	***
Ireg7_2	0.0314	0.6		Ireg7_2	0.0693	0.7	
Ireg7_3	0.0485	0.8		Ireg7_3	0.0445	0.6	
Ireg7_4	0.1373	2.2	**	Ireg7_4	0.1460	1.9	*
Ireg7_5	0.1708	2.1	**	lreg7_5	variable omitted		
Ireg7_6	0.5424	9.4	***	Ireg7_6	0.4151	5.5	***
Ireg7_7	0.3011	5.1	***	Ireg7_7	0.1895	2.1	**
_cons	7.5327	108.7	***	_cons	7.7538	64.7	***

Source: Regression analysis of 1998 Viet Nam Living Standards Survey.
Note: The dependent variable is log of per capita expenditure.
* coefficient is significant at the 10% level, ** at the 5% level, and *** at the 1% level.

showing that after controlling for differences in endowments (education, housing characteristics, and ownership of consumer durables), differences in per capita expenditure between ethnic minority households and others remain, but are much smaller.

3.2 Education

In both urban and rural areas, the level of schooling of the head of household is a good predictor of a household's per capita expenditure.[7] The five dummy variables that represent the education of the head are jointly significant at the 1 percent level in both rural and urban areas (see Table 3). In rural areas, heads of household who complete primary school earn 6 percent more than those not completing primary school. In urban areas, households whose head has completed primary or lower secondary school do not seem to be better off than those whose head has not completed primary school, but higher levels of education are associated with significantly higher earnings (see Table 2).

In general, the educational level of the spouse is less significant than that of the household head as a predictor of per capita expenditure.[8] In the rural areas, only the highest two levels of education of the spouse (advanced technical training and post-secondary education) show any significant effect relative to the base level (not completing primary school). The education of the spouse is a better predictor in urban areas than in rural areas (see Table 2).

3.3 Occupation

The occupation of the head of household is a statistically significant predictor of per capita expenditure in rural and urban areas.[9] In rural areas, the first three

---

[7] We also experimented with using the number of years of the education for the household and spouse as explanatory variables, but found that the level of education completed gave better results.

[8] Education of the spouse may have other benefits, such as improved health or nutrition, that are not captured by the measure of welfare used in this analysis, per capita expenditure. Note that 11.4 per cent of spouses in the VLSS are male.

[9] Although information on the employer of households heads is available in both the Census and the VLSS, the categories they use to describe different categories of employers differ substantially and cannot be reconciled.

Table 3. Tests of significance of groups of explanatory variables in urban-rural regressions

Sector	Variables	df1	df2	F statistic	Probability
Rural	Education of head of household	5	129	7.80	0.0000 ***
	Education of spouse	6	129	1.97	0.0738 *
	Occupation of head	6	129	12.65	0.0000 ***
	Type of housing	2	129	14.00	0.0000 ***
	Main source of water	2	129	9.69	0.0001 ***
	Type of sanitary facility	2	129	15.64	0.0000 ***
	Region	6	129	26.20	0.0000 ***
Urban	Education of head of household	5	55	4.01	0.0036 ***
	Education of spouse	6	55	3.10	0.0110 **
	Occupation of head	6	55	2.90	0.0157 **
	Type of housing	2	55	10.76	0.0001 ***
	Main source of water	2	55	17.17	0.0000 ***
	Type of sanitary facility	2	55	4.12	0.0216 **
	Region	5	55	10.29	0.0000 ***

Source: Regression analysis of per capita expenditure using 1998 VLSS

Note: The dependent variable is log of per capita expenditure.

* coefficient is significant at the 10% level, ** at the 5% level, and *** at the 1% level.

occupational categories (political leaders/managers, professionals/technicians, and clerks/service workers) are significantly better off than households in which the head is not working. On the other hand, there is no statistically significant difference between the expenditure of farm households and households with non-working heads (see Table 2). This somewhat counter-intuitive finding probably reflects the fact that non-working heads include retirees as well as a disproportionate number of urban workers who can "afford" to look for work.

In urban areas, households whose head is a leader/manager are significantly better off than those with non-working heads, while those whose head is an unskilled worker are significantly worse

For this reason, a set of dummies for employer of the household head were not included in the predictive regressions.

off (see Table 2). This suggests that in urban areas, a non-working head of household is not a reliable indicator that the household is poor.

3.4 Housing and basic services

Various housing characteristics are good predictors of expenditures. Living in a house or other dwelling made of permanent rather than temporary materials is associated with 19 percent (24 percent) higher per capita expenditure in rural (urban) areas.[10] Similarly, having a house of semi-permanent rather than temporary materials implies a significantly higher level of per capita expenditure. The living area of houses is also a useful predictor of household well being. Houses in Vietnam have an average living area of about 45 square meters, and each 10 percent increase in area is associated with a 12-30 percent increase in per capita expenditure, depending on the area of residence (urban or rural) and the type of house (permanent or semi-permanent)[11].

Electrification[12] is a statistically significant predictor of household welfare in rural areas, where 71 percent of the household have access to electricity. By contrast, in urban areas, where 98 percent of the households are already electrified, electricity is not a significant predictor of expenditures (see Table 2).

The main source of water is also useful in distinguishing poor households. In rural areas, households with access to well water have higher level of per capita expenditures than households using river or lake water (the omitted category). Access to tap water is not a statistically significant predictor of expenditures in rural areas, presumably because just 2 percent of the rural households fall into this category. By contrast, in urban areas more than half the sample households (58 percent) have access to tap water, and this variable is a good predictor of urban per capita expenditures.

---

10 Because the permanent housing dummy enters both as a separate variable (Ihouse_1) and in the interaction term htypla1 (=Ihouse_1×ln(area)), the marginal effect is calculated as $\beta_{Ihouse\_1} + \beta_{htypla1} \times \ln(\text{area})$. We evaluate the marginal effect at the mean values of ln(area), which are 3.72 in rural areas and 3.66 in urban areas.

11 The Census did not collect information on the area of houses made of temporary materials, so we cannot use housing area to help predict expenditures for these houses.

12 More specifically, this variable refers to whether the household said that electricity was the main source of lighting for the house.

Finally, sanitation facilities can be used to separate poor from non-poor households. In rural areas, flush toilets and latrines are statistically significant indicators of higher per capita expenditure at the 5 percent level. In urban areas, having a flush toilet is a significant predictor of expenditures at the 5 percent level but having a latrine is not (see Tables 2).

3.5 Consumer durables

Television ownership is one of the strongest predictors of per capita expenditures, being a statistically significant predictor in both urban and rural areas. Radio ownership is almost as good a predictor, being statistically significant at the 1 percent level in both urban and rural areas. As expected, the coefficient for radio ownership is smaller than that of television ownership (see Table 2). In Section 5 below, we examine to what extent the addition of variables reflecting ownership of consumer durables or housing characteristics can improve the geographic targeting of the poor.

3.6 Region

Regional dummy variables were included in the urban and rural regression models, with, the Northern Uplands, as the base region. Even after controlling for other household characteristics, rural households in the four southern regions are shown to be better off than those in the Northern Uplands. The coefficient in the Southeast is the largest, implying that households in this region have expenditure levels 72 percent higher than similar households in the Northern Uplands. A similar pattern holds for urban households (see Table 2). The regional dummy variables are jointly significant at the 1 percent level in both urban and rural areas (see Table 3).

4. Poverty Maps of Vietnam

As discussed in Section 2.3, the second stage in constructing a poverty map is to combine the regression coefficients estimated from the VLSS in the first stage and the Census data on the same household characteristics. This gives us predicted expenditures for each household in the Census which are then used to estimate the incidence of poverty (the poverty headcount) for individual regions and provinces, as well as the standard errors associated with these estimates. We present the estimates of the incidence of poverty first at the regional level and then at the provincial level.

4.1 Regional poverty estimates

Regional poverty headcounts and their standard errors, as estimated directly from the 1998 Vietnam Living Standards Survey, are shown in the first two columns of Table 4. For the country as a whole, the incidence of poverty is 37.4 percent with a 95 percent confidence interval of ± 3.2 percentage points. The regional poverty headcounts range from 0.9 percent in urban Hanoi and Ho Chi Minh City to 65.2 percent in the rural Northern Uplands. The standard errors suggest that the degree of precision in the estimates of regional poverty using the VLSS is relatively low: four of the nine regions have confidence limits of ± 10 percentage points or more.

By combining the urban-rural regression models and the Census data (as described in Section 2), we get an alternative set of estimates of regional headcount poverty rates and standard errors, shown in the second pair of columns in Table 4. Seven of the nine regional estimates are within 3 percentage points of the corresponding estimate from the VLSS. However, the Census-based poverty estimates tend to be less extreme: they are higher than

Table 4. Comparison of original and Census-based poverty headcounts

Region	VLSS 1998		Urban-rural regressions with Census data		Stratum regression with Census data	
	Poverty	Std error	Poverty	Std error	Poverty	Std. Error
Hanoi & HCMC	0.009	0.004	0.037	0.007	0.031	0.009
Other urban	0.138	0.021	0.145	0.012	0.146	0.014
Rural N Uplands	0.652	0.057	0.598	0.011	0.626	0.037
Rural Red R Delta	0.361	0.038	0.379	0.006	0.407	0.031
Rural N C Coast	0.488	0.058	0.513	0.011	0.490	0.036
Rural S C Coast	0.436	0.075	0.460	0.010	0.400	0.028
Rural C Highlands	0.524	0.097	0.533	0.016	0.525	0.046
Rural Southeast	0.130	0.022	0.234	0.004	0.173	0.018
Rural Mekong Delta	0.412	0.033	0.397	0.007	0.386	0.031
Total	0.374	0.016	0.365	0.012	0.365	0.011

Source: Data from 1998 VLSS and 3% sample of 1999 Population and Housing Census

Note: Poverty headcounts are expressed as fractions rather than percentages.

the VLSS estimates where the incidence of poverty is low (such as in the rural Southeast and in urban areas) and lower where the incidence is high (such as in the rural Northern Uplands). In every region except one (Hanoi and Ho Chi Minh City), the standard errors of the Census based estimates are substantially smaller than those of the VLSS estimates. Apparently, the gains in accuracy from using a larger sample exceed the losses due to estimating expenditure based on household characteristics.

According to the urban-rural regression results in Table 4, the rural Northern Uplands is the poorest region. In fact, it is significantly poorer than the other eight regions at the 1 percent confidence level (see Table 5). The rural Central Highlands and the rural North Central Coast are the next poorest regions, although there is no statistically significant difference between the two. Then follows the rural South Central Coast, the rural Mekong Delta, and the rural Red

Table 5. Differences in regional poverty headcounts and their statistical significance

Region	Hanoi & HCMC	Other urban	Rural N Uplands	Rural Red R Delta	Rural N C Coast	Rural S C Coast	Rural Highlands	Rural Southeast
Other urban	-0.109 ***	-						
	(0.012)	-						
Rural N Uplands	-0.561 ***	-0.452 ***	-					
	(0.013)	(0.016)	-					
Rural Red R Delta	-0.343 ***	-0.234 ***	0.218 ***	-				
	(0.009)	(0.013)	(0.012)	-				
Rural N C Coast	-0.477 ***	-0.368 ***	0.084 ***	-0.134 ***	-			
	(0.013)	(0.016)	(0.016)	(0.013)	-			
Rural S C Coast	-0.438 ***	-0.330 ***	0.123 ***	-0.096 ***	0.038 ***	-		
	(0.012)	(0.015)	(0.016)	(0.012)	(0.014)	-		
Rural C Highlands	-0.481 ***	-0.372 ***	0.081 ***	-0.138 ***	-0.004 n.s.	-0.042 **	-	
	(0.017)	(0.020)	(0.021)	(0.018)	(0.020)	(0.019)	-	
Rural Southeast	-0.089 ***	0.020 n.s.	0.472 ***	0.254 ***	0.388 ***	0.349 ***	0.392 ***	-
	(0.008)	(0.012)	(0.012)	(0.007)	(0.012)	(0.011)	(0.017)	-
Rural Mekong Delta	-0.360 ***	-0.252 ***	0.201 ***	-0.017 **	0.117 ***	0.078 ***	0.120 ***	-0.271 ***
	(0.009)	(0.013)	(0.013)	(0.008)	(0.013)	(0.012)	(0.017)	(0.008)

Source: Data from 1998 VLSS and 3% sample of 1999 Population and Housing Census

Note: Differences expressed as poverty headcount of column region minus poverty headcount of row region. Standard errors in parentheses.

* statistically significant at the 10% level, ** at the 5% level, and *** at the 1% level.

n.s. not statistically significant at the 10% level

River Delta, the differences being statistically significant in each case. The rural Southeast and "Other urban" areas are significantly less poor than the rural Red River Delta, but the difference between the two is not statistically significant. The ninth region, Hanoi and Ho Chi Minh City, is significantly less poor than any of the other eight regions (see Tables 4 and 5).

Combining the *stratum-level* regression models with the Census data yields results similar to those based on the urban-rural regression models, as shown in the last two columns of Table 4. Again, the poverty estimates are less extreme than the VLSS estimates and the standard errors are somewhat lower. One notable difference is that the standard errors of the poverty estimates based on the stratum-level regression models are higher, often two to three times higher, than those based on the urban-rural regression models.

4.2 Provincial poverty estimates

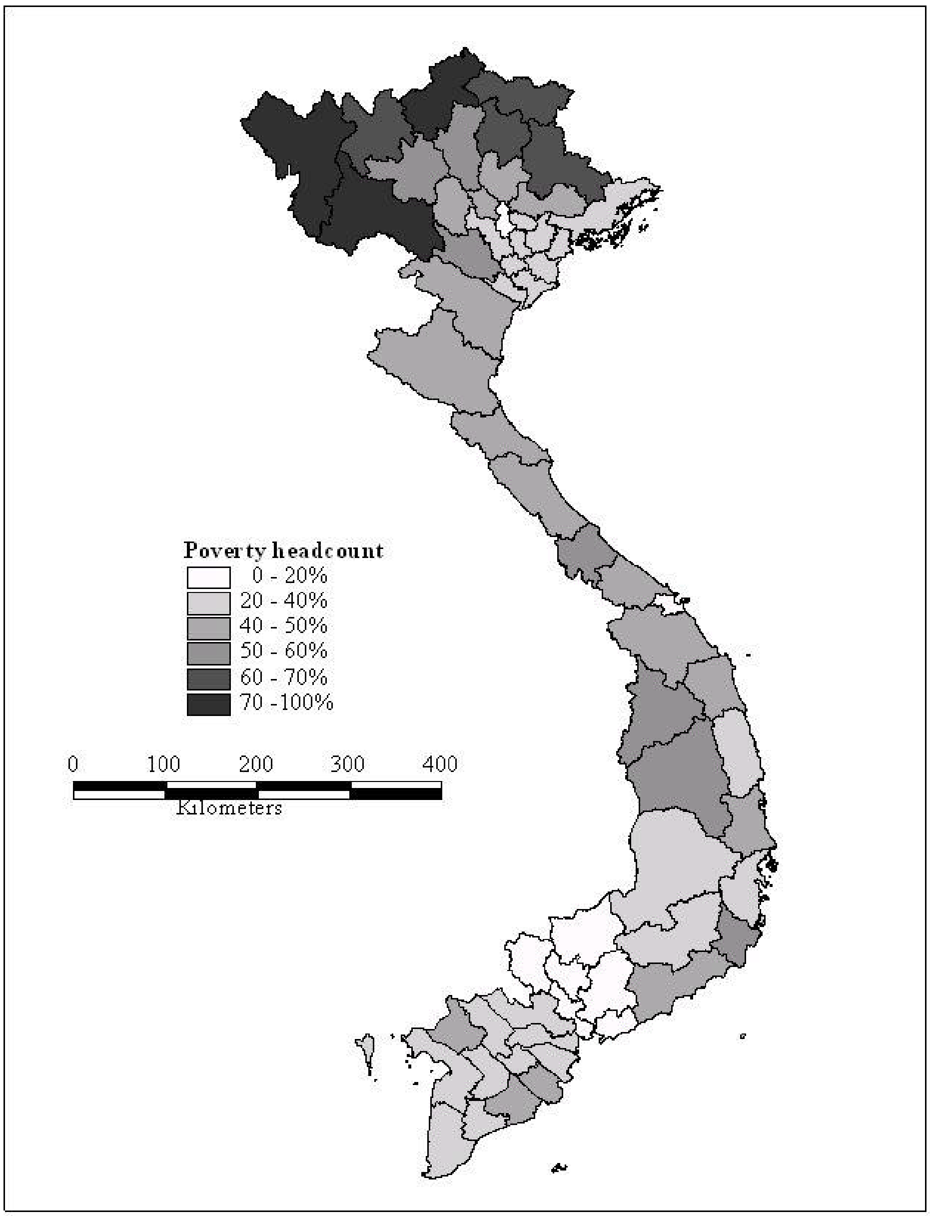
One of the main advantages of using Census data is that they allow us to generate reliable estimates of poverty for smaller geographic units, such as provinces or districts, which would be difficult or impossible to estimate with a household sample survey such as the VLSS[13]. Tables 6 shows the estimated provincial poverty rates, along with the standard errors of the estimates, based on the urban-rural regression models (the corresponding results from the stratum-level regressions are given in Annex 2). Figure 1 shows the geographic distribution of poverty at the provincial level, also based on the urban-rural regression models.

The results indicate that Lai Chau, located at the extreme northwest corner of Vietnam, is the poorest province, with over three-quarters of its population living below the poverty line. The next five poorest provinces (Ha Giang, Son La, Cao Bang, Lao Cai, and Lang Son) are all provinces in the Northern Uplands on the northern border with China or the western border with

13 There are three factors that complicate using the VLSS for estimating provincial poverty. First, three provinces are not included in the VLSS sample. Second, in the remaining provinces, the sample size is small: most provinces have less than 100 households and some have as few as 32. Third, the sample (and hence the sampling weights) are not designed to produce precise estimates at the provincial level. For example, the proportion of urban households in each province is not accurate, even after applying sampling weights.

Laos. In fact, the ten poorest provinces are all in the Northern Uplands. This is probably a reflection of their mountainous topography, distance from major markets, and limited

Figure 1. Incidence of poverty by province



Source: Estimated from urban-rural regression models of 1998 VLSS and household characteristics in the 1999 Population and Housing Census infrastructure, all of which reduce the returns to agriculture in this region. Ethnic minorities also comprise more than half of the population of these provinces.

Poverty is not limited to the Northern Uplands, however. The North Central Coast comprises six provinces, all of which are among the poorest 21 provinces in the country. The incidence of poverty in these provinces ranges from 44 percent to 52 percent.

The Central Highlands region includes three provinces. Two of the three, Kon Tum and Gia Lai, are among the 15 poorest provinces in Vietnam, with poverty headcounts of more than 50 percent. The third province, Dak Lak, is more prosperous, with a poverty headcount similar to the national average. This is probably due to the importance of coffee production. Vietnam now exports US$ 500 million of coffee per annum, most of which is grown in Dak Lak province.

Poverty is less severe in the southern regions, although each region has at least one province with a poverty headcount over 40 percent. The Southeast region is the least poor region, but it has two provinces, Ninh Tuan and Binh Tuan, with poverty headcounts over 40 percent. These provinces are farther from Ho Chi Minh City than the other provinces in the Southeast. In the South Central Coast, Quang Ngai has a poverty headcount of 47 percent. In the Mekong River Delta, Soc Trang, Tra Vinh, and An Giang have rates over 40 percent.

The lowest incidence of poverty is found in Ho Chi Minh City (less than 5 percent), followed by four provinces in the Southeast (Binh Duong, Ba Ria-Vung Tau, Dong Nai, and Tay Ninh) all of which have poverty headcounts under 15 percent. The headcounts for Hanoi and Da Nang are both close to 15 percent.

Poverty headcounts in *rural* areas are similar to the overall provincial poverty levels, which is not surprising given the large proportion of the population living in rural areas in most provinces (see Table 6 and Figure 2). Rural poverty is greatest in the border provinces of the Northern Uplands. The Central Highlands provinces of Gia Lai and Kon Tum are among the ten poorest provinces in terms of rural poverty.

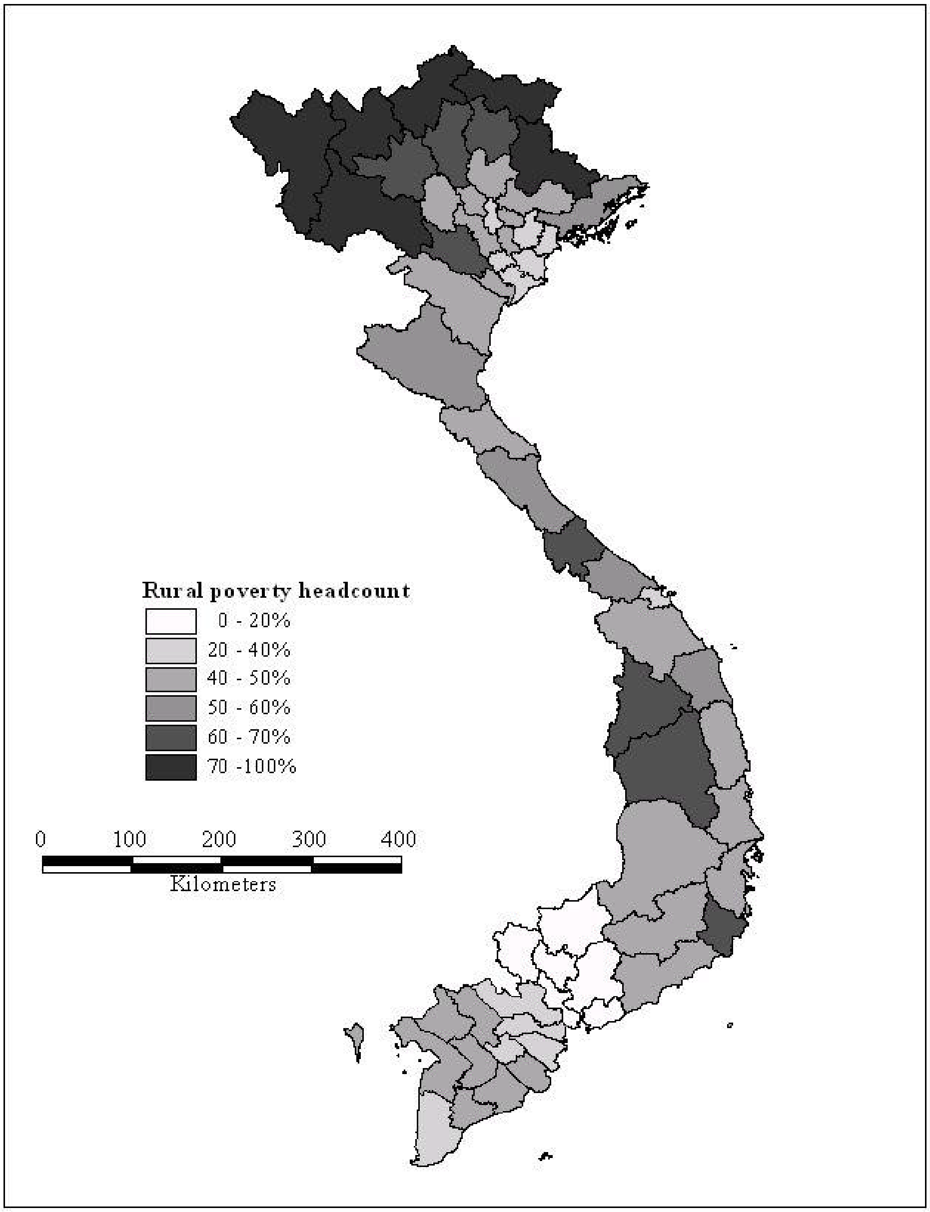
Table 6. Provincial poverty headcounts estimated with urban-rural regression model

			Poverty headcount			Standard errors		
Rank	Province	Region	Rural	Urban	Total	Rural	Urban	Total
1	Lai Chau	NU	0.857	0.221	0.777	0.038	0.036	0.034
2	Ha Giang	NU	0.770	0.195	0.722	0.039	0.032	0.036
3	Son La	NU	0.795	0.153	0.714	0.039	0.029	0.034
4	Cao Bang	NU	0.739	0.142	0.675	0.037	0.034	0.033
5	Lao Cai	NU	0.747	0.197	0.652	0.043	0.031	0.036
6	Lang Son	NU	0.724	0.141	0.617	0.038	0.033	0.032
7	Bac Kan	NU	0.676	0.189	0.609	0.039	0.037	0.034
8	Hoa Binh	NU	0.655	0.155	0.586	0.041	0.028	0.036
9	Tuyen Quang	NU	0.635	0.161	0.583	0.043	0.026	0.038
10	Yen Bai	NU	0.644	0.165	0.550	0.044	0.027	0.036
11	Gia Lai	CH	0.650	0.194	0.538	0.062	0.032	0.047
12	Ninh Thuan	SE	0.618	0.214	0.525	0.041	0.038	0.033
13	Kon Tum	CH	0.670	0.221	0.522	0.061	0.035	0.043
14	Quang Tri	NCC	0.618	0.192	0.520	0.043	0.034	0.034
15	Quang Binh	NCC	0.532	0.132	0.491	0.044	0.028	0.040
16	Nghe An	NCC	0.515	0.140	0.477	0.046	0.029	0.041
17	Quang Ngai	SCC	0.513	0.153	0.474	0.043	0.030	0.038
18	Thua Thien - Hue	NCC	0.579	0.185	0.472	0.043	0.033	0.033
19	Bac Giang	NU	0.494	0.164	0.470	0.050	0.028	0.047
20	Thanh Hoa	NCC	0.492	0.135	0.460	0.045	0.027	0.041
21	Ha Tinh	NCC	0.474	0.151	0.445	0.044	0.030	0.040
22	Vinh Phuc	NU	0.470	0.199	0.442	0.052	0.032	0.047
23	Binh Thuan	SE	0.498	0.235	0.435	0.041	0.040	0.033
24	Phu Tho	NU	0.482	0.132	0.431	0.049	0.024	0.042
25	Soc Trang	MRD	0.463	0.244	0.424	0.034	0.040	0.029
26	Thai Nguyen	NU	0.495	0.126	0.419	0.047	0.023	0.038
27	Tra Vinh	MRD	0.452	0.191	0.418	0.034	0.032	0.030
28	Phu Yen	SCC	0.469	0.188	0.416	0.042	0.036	0.035
29	Quang Nam	SCC	0.443	0.191	0.408	0.041	0.035	0.036
30	An Giang	MRD	0.454	0.196	0.406	0.033	0.036	0.027
31	Dac Lac	CH	0.451	0.176	0.395	0.063	0.029	0.050
32	Ha Tay	RRD	0.417	0.125	0.395	0.033	0.028	0.031
33	Dong Thap	MRD	0.424	0.195	0.391	0.032	0.036	0.028
34	Binh Dinh	SCC	0.460	0.179	0.391	0.041	0.033	0.032
35	Ninh Binh	RRD	0.424	0.109	0.385	0.033	0.026	0.029
36	Bac Ninh	NU	0.405	0.166	0.383	0.050	0.028	0.046
37	Hung Yen	RRD	0.403	0.163	0.383	0.032	0.037	0.030
38	Kien Giang	MRD	0.428	0.210	0.380	0.034	0.036	0.028
39	Bac Lieu	MRD	0.430	0.207	0.377	0.033	0.037	0.027
40	Ha Nam	RRD	0.391	0.143	0.376	0.033	0.031	0.031
41	Quang Ninh	NU	0.519	0.155	0.357	0.048	0.026	0.029
42	Nam Dinh	RRD	0.385	0.110	0.351	0.032	0.026	0.028
43	Can Tho	MRD	0.402	0.156	0.349	0.031	0.031	0.025
44	Ca Mau	MRD	0.388	0.152	0.345	0.032	0.030	0.027
45	Lam Dong	SE	0.458	0.144	0.337	0.061	0.024	0.039
46	Vinh Long	MRD	0.360	0.148	0.330	0.031	0.030	0.027
47	Thai Binh	RRD	0.345	0.075	0.330	0.033	0.021	0.032
48	Ben Tre	MRD	0.342	0.137	0.325	0.031	0.029	0.028
49	Hai Duong	RRD	0.353	0.106	0.319	0.032	0.027	0.028
50	Khanh Hoa	SCC	0.416	0.126	0.311	0.040	0.024	0.027
51	Long An	MRD	0.335	0.151	0.305	0.031	0.031	0.027
52	Hai Phong	RRD	0.395	0.074	0.286	0.032	0.019	0.022
53	Tien Giang	MRD	0.301	0.105	0.276	0.030	0.025	0.026
54	Binh Phuoc	SE	0.197	0.076	0.179	0.028	0.017	0.024
55	Da Nang	SCC	0.346	0.106	0.156	0.038	0.022	0.019
56	Ha Noi	RRD	0.306	0.037	0.152	0.031	0.010	0.015
57	Tay Ninh	SE	0.130	0.081	0.124	0.019	0.017	0.016
58	Dong Nai	SE	0.137	0.048	0.111	0.020	0.011	0.014
59	Ba Ria-Vung Tau	SE	0.109	0.062	0.090	0.016	0.013	0.011
60	Binh Duong	SE	0.092	0.051	0.079	0.014	0.012	0.010
61	TP Ho Chi Minh	SE	0.082	0.036	0.044	0.014	0.008	0.007
	Total		0.441	0.111	0.365	0.015	0.011	0.012

Source: Estimated from 1998 VLSS and 3% sample of 1999 Population and Housing Census

Note: A poverty headcount of 0.406 for An Giang implies that 40.6 percent of the population in An Giang live in households with per capita expenditures below the 1998 GSO/WB poverty line. The region codes are NU=Northern Uplands, RRD=Red River Delta, NCC=North Central Coast, SCC=South Central Coast, CH=Central Highlands, SE=Southeast, and MRD=Mekong River Delta.

Figure 2. Incidence of rural poverty by province



Source: Estimated from urban-rural regression models of 1998 VLSS and household characteristics in the 1999 Population and Housing Census

As expected, the incidence of poverty in urban areas is consistently lower than that in rural areas. Even in the poorest provinces, where over 70 percent of the rural population are poor, urban poverty is below 25 percent. In contrast, the difference between rural-urban poverty headcounts is relatively small in the more prosperous provinces in the Southeast (see Table 6).

In order to determine whether the poverty estimates for any two provinces are statistically different from one another, the standard error of the difference between their poverty headcounts must be calculated. This statistic can be computed using the equations described in Section 2.3, which take into account the modeling error, the idiosyncratic error, and the sampling error associated with the 3% sample Census. We have calculated the standard errors of these differences (based on the urban-rural regressions) for the 1830 possible pairs of provinces, and their rural and urban sub-samples, together with the 61 urban-rural pairs in the same province. Our results can be summarized as follows:

- About one-quarter (23 percent) of the provincial pairs with a 6 percentage point gap[14] in their poverty incidence are significantly different from each other at the 5 percent level of statistical significance. Forty-three percent of the provincial pairs with an 8 percentage point gap and 70 percent of those with a 10 percentage point gap have statistically different poverty levels. This implies that poverty headcounts are generally not statistically different from one another in provinces that are adjacent to each other in poverty rankings. Provinces that are four to five provinces away from each other in the ranking, however, will usually have statistically significant differences in their poverty headcounts.

- Poverty headcounts in 65 percent of the rural provincial pairs are significantly different from one another (at the 5 percent level), but just 33 percent of the urban

[14] A 6 percent gap refers to a gap greater than 5.5 percent and less than or equal to 6.5 percent.

pairs are. This is largely due to the fact that rural areas have higher and more diverse poverty headcounts, so the (absolute) differences are larger.

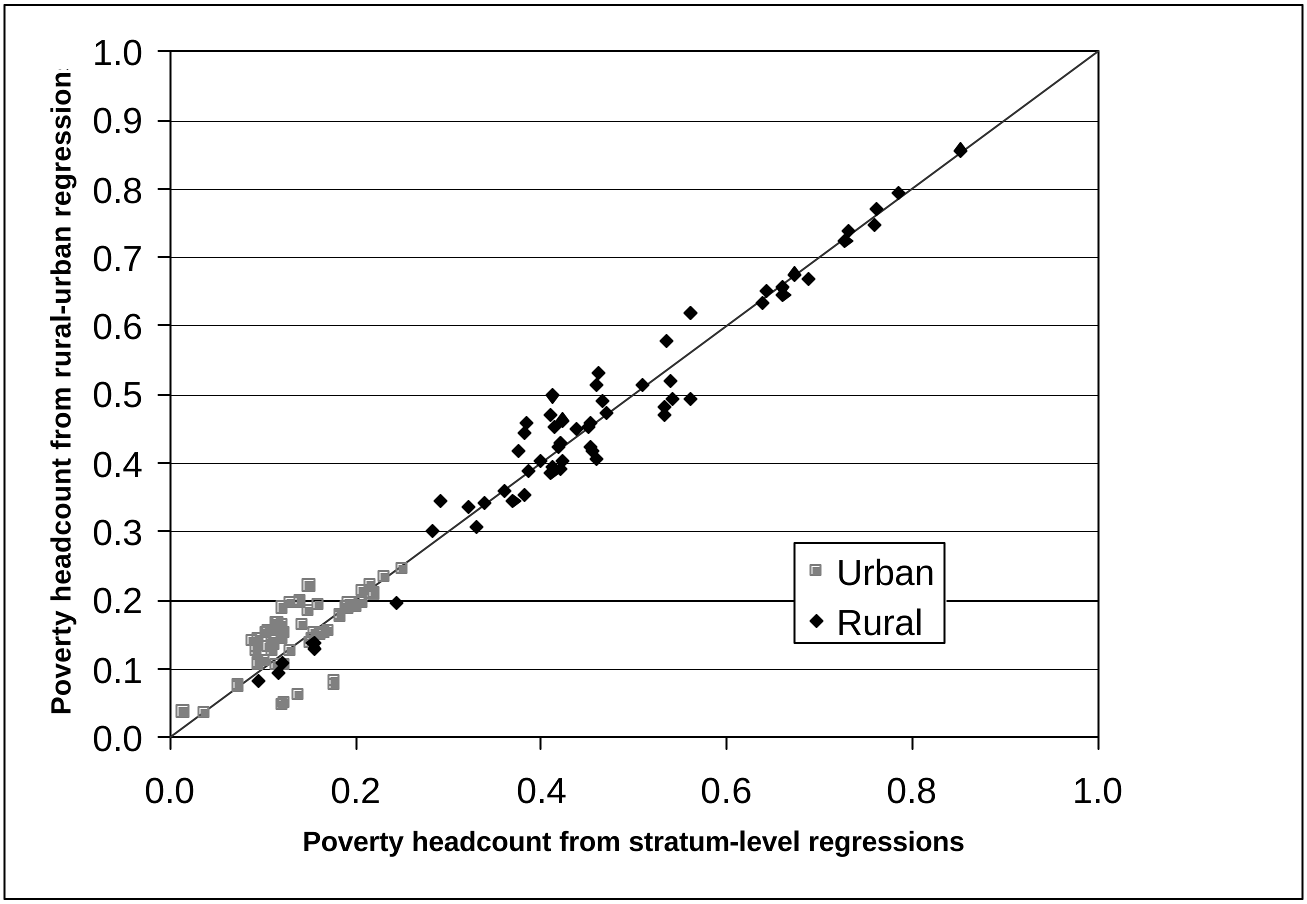
- In every province except one (Tay Ninh), the incidence of rural poverty is significantly higher than that of urban poverty.

Finally, we examined the sensitivity of these results to the type of regression models estimated in the first step of our analysis. In particular, how different are the results when we use stratum-level regressions instead of urban-rural regressions? The average (absolute value) gap between provincial poverty headcounts obtained from these two regression models is 2.2 percentage points. Just eight provinces have differences of more than 5 percentage points and none have differences of more than 10 percentage points. Furthermore, the ranking of the ten poorest provinces is the same according to the two approaches.

Figure 3 shows the similarity of the rural and urban poverty headcounts for each province (identical headcounts would be represented by points along the diagonal line). The two methods are most similar for the poorest rural regions, where the difference in estimates is typically just 1 percentage point. They are less similar for more prosperous rural areas and for urban areas. The urban poverty headcounts often differ by four to eight percentage points.

We also compared the standard errors of the provincial poverty headcounts. Those based on the urban-rural regression models were often (72 percent of the time) lower than the corresponding standard errors based on the stratum-level regressions. For the poorest provinces, the standard errors of the headcount based on the stratum level regressions are roughly twice as large as those based on the urban-rural regressions.

Figure 3. Provincial Poverty Headcounts estimated using Urban-Rural and Stratum-Level Regression Models



5. The Potential of Geographic and Additional Targeting Variables

Given knowledge about where poor people/households live, a natural question to ask is how effective geographical variables are in identifying the poor. Experience in other countries indicates that the ability to target poor households typically improves with greater geographical disaggregation (Baker and Grosh, 1984; Bigman and Fofack, 2000). Since many of Vietnam's anti-poverty programs use highly disaggregated listings of "poor and remote communes", one would expect the efficiency of its geographically targeting programs to be quite high.[15] It would also be interesting to know whether the poor can be identified more accurately if additional information other than place of residence is available. Implicitly, this is what the commune/district level staff of Vietnamese Ministry of Labor, Invalids and Social Assistance does when determining whether a household is classified as poor. If a household is classified as poor, it is eligible to receive various benefits such as health cards, free or subsidized primary schooling for children, and sometimes exemption from local taxes. Put differently, can the geographic targeting of transfers to the poor be improved by the use of the type of additional socio-economic variables that can be collected easily in a "quick-and-dirty" enumeration of households?

We assess the efficacy of different targeting variables using a relatively novel technique: Receiver Operating Characteristic (ROC) curves, a graphic way of portraying the accuracy of a diagnostics test originally developed for use in electrical engineering and signal-processing (Stata Corporation, 2001a). An ROC curve shows the ability of a diagnostic test to correctly distinguish between two states or conditions. In the context of poverty targeting, an ROC curve plots the probability of a test correctly classifying a poor person as poor (known as the test's "sensitivity") on the vertical axis against 1 minus the probability of the same test correctly classifying a non-poor person as non-poor on the horizontal access (known as the test's "specificity").[16] When the diagnostic test (here the

15 Vietnam has two official lists for identifying poor and remote communes. The Ministry of Labour, Invalids and Social Assistance (MOLISA) maintains a list of "poor communes", most of which are located in coastal and lowland areas under Programme 133. In addition, the Committee for Ethnic Minorities in Mountainous Areas (CEMMA) is responsible for identifying "especially difficult mountainous and remote communes" under Programme 135. Since the geographic areas in which these two programmes operate are reasonably distinct, we have combined the two into one list of "poor and remote communes" for our analysis of targeting. This list was then matched to commune information in the VLSS to identify households living in areas identified as poor by MOLISA or CEMMA.

16 ROC curves can be linked to the occurrence of Type I and Type II errors familiar from conventional statistical hypothesis testing (known as "false positives" and "false negatives" in epidemilogy and medicine and F and E errors in the targeting literature) as follows. Sensitivity is 1 minus the probability of a Type I error (incorrectly classifying a poor

values of a targeting variable) can take several discrete values, the ROC curves will consist of a series of linear segments corresponding to these discrete values. The greater the area under an ROC curve and the closer it is to the left-hand side vertical and top horizontal axes, the greater is the efficacy of a diagnostic test. The closer a ROC curve is to the 45-degree line, the weaker is its efficacy.

To our knowledge, the only previous use of ROC analysis for analyzing the impact of poverty targeting is by Wodon (1997) using household survey data from Bangladesh. As Wodon points out, unlike conventional statistical hypothesis tests ROC analysis can take account of continuous as well as categorical targeting variables. However, like conventional hypothesis tests, ROC analysis can only be employed for dichotomous outcome variables (so that it can be used for the conventional poverty headcount but not for higher-order poverty measures such as the poverty-gap and squared poverty gap).

Figure 4 shows an example of two pairs of ROC curves drawn using data from 1998 VLSS. Since the curve for the index of radio and television ownership in rural areas lies everywhere above and to the left of the curve for the education level completed by the household head, Panel (a) shows that use of the television and radio ownership variables unambiguously dominates that for education of the household head as a targeting variable. Note that the ROC for the index of radio and television ownership has four linear segments corresponding to the four values of the index, while the ROC curve for the head's education has six segments corresponding to the six educational levels a household head may complete. Panel (b) shows the contrasting situation in which the ROC for quintiles of land area and the number of children per household cross, in which case neither variable unambiguously dominants the other from a targeting perspective.[17] Of course, it will also usually be the case that some combination (linear or otherwise) of the two variables will further improve the efficacy of a test.

---

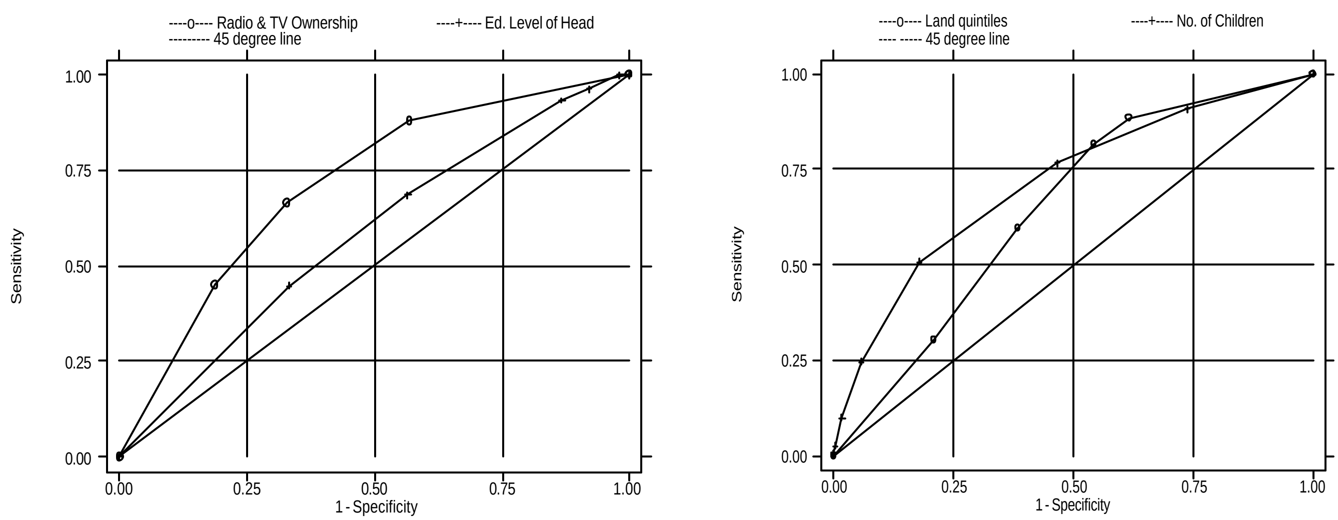
households as non-poor) while 1 minus the specificity of a test is the same as the probability of a Type II error (incorrectly classifying a non-poor household as poor). In many respects this is akin to describing whether "a glass is half-empty or half-full", in that both are simply different methods of presenting the same data.

[17] This is rather similar to the problems encountered in making unambiguous comparisons of inequality when the Lorenz curves cross or in making comparisons of inequality when cumulative income distribution curves cross.

As long as a potential targeting variable increases in value as the likelihood of poverty increases (i.e., it is “monotonically increasing with the risk of failure”), then the area under an ROC curve can be used for ranking the efficacy of different targeting variables (Stata Corporation, 2001a). The more a test’s ROC curve is bowed toward the upper left-hand corner of the graph, the greater is the accuracy of the test. Since the ROC curves are bounded by the interval [0,1], the maximum value for the area under an ROC curve is 1.0 (in which case the test would predict poverty perfectly and the ROC curve would coincide with the left-hand vertical and top horizontal axes). In contrast, a test with no predictive power would correspond to an area of 0.5 under the ROC curve (which would itself coincide with the 45-degree line in the ROC diagram). Table 7 shows the

Figure 4. Receiver Operating Characteristic Curves for Selected Targeting Variables

(a) (b)



area under the ROC curves for a number of possible additional targeting variables that the information would be obtained relatively easily in a "quick and dirty” survey.

It can be seen that the current system for classifying “poor and remote communes” does not perform particularly well in identifying poor people, especially for the “overall” poverty line. Although the poor and remote communes list has a relatively low probability (7.7 percent) of incorrectly identifying a non-poor person as poor, it has an high probability (80.5 percent) of classifying a poor person as non-poor – for the simple reason that the vast majority of poor people in Vietnam do not live in an officially designated poor or remote commune. With the exception of educational level of the spouse, land allocated and livestock owned in rural areas, Table 7 shows that household level targeting variables are generally much better at identifying poor individuals than whether or not they live in a poor and remote commune. The four categories of provincial poverty headcounts identified in our national poverty map also do quite well according to this criterion. Nonetheless, as shown by this and the ranking of poor communes according to their mean expenditures, there is considerable potential for improving the targeting of Vietnam's poor and remote communes programs.

Table 7 also shows that the most effective poverty targeting variables are ones related to housing quality and ownership of durable assets. Floor type is generally a better predictor of both food poverty and overall poverty than roof or toilet type.[18] The level of education completed by household heads and their spouses performs considerably better as a targeting indicator in urban than in rural areas. Demographics, as proxied by the number of children under 15 years of age (the age by which Vietnam children should have completed lower secondary school) are a better indicator of food poverty than overall poverty in both rural and urban areas. Ethnicity of the household head is a reasonable predictor of both food and overall poverty in rural areas, but performs poorly in urban areas where few ethnic minority households live.

An unexpected result is that a simple index of radio and television ownership is a better targeting indicator than all other asset, demographic or educational variables. Indeed, inspection of Table 7 will confirm that the radio and television ownership index dominates all other targeting variables with the exception of communes ranked by the level of their median per capita expenditures. Using a cut-off point corresponding to ownership of neither a radio nor a television, the index is able to correctly classify some 76 percent of poor people in the VLSS sample.[19]

---

[18] Ownership of the dwelling in which a household lives was considered for inclusion in the list of asset based targeting variables, but found to perform poorly because the vast majority of households in the VLSS98 sample (5703 out of 5999) own their own dwellings.

[19] It may seem surprising that in a country with Vietnam's level of per capita income, radio and television ownership has such potential for targeting the poor. Radio and television ownership is however, quite widespread throughout Vietnam with 53 percent of households owning a television and 45 percent of households owning a radio according to the 1999 Population and Housing Census.[19] Many of the televisions owned, especially in rural areas, are relatively inexpensive 14 inch, battery operated televisions produced in China. Of course, the use of an index of television and radio ownership for targeting would be problematic, as it would be relatively easy for households to

Table 7. Accuracy of different variables in targeting poor households

	Targeting accuracy (area under ROC curve)					
	Rural		Urban		All Vietnam	
Targeting Variable	Food Poverty	Overall Poverty	Food Poverty	Overall Poverty	Food Poverty	Overall Poverty
Poor or Remote Commune	0.585	0.559	0.554	0.520	0.589	0.559
Categories in National Poverty Map	0.641	0.622	0.620	0.645	0.663	0.650
Communes ranked by median expenditure	0.829	0.790	0.726	0.808	0.849	0.827
Land allocated (quintiles)	0.529	0.542	n/a	n/a	0.619	0.646
Livestock owned (animal eq. units)	0.474	0.448	0.591	0.541	0.467	0.441
Educational Level of Household Head	0.601	0.579	0.715	0.685	0.625	0.609
Educational Level of Spouse *	0.570	0.554	0.739	0.727	0.602	0.597
Number of Children under 15	0.733	0.690	0.753	0.789	0.742	0.714
Number of Females	0.636	0.618	0.578	0.671	0.632	0.616
Ethnicity	0.642	0.612	0.495	0.500	0.649	0.614
Floor Type	0.696	0.665	0.694	0.773	0.734	0.720
Roof Type	0.630	0.585	0.687	0.658	0.637	0.594
Toilet Type	0.597	0.577	0.773	0.730	0.650	0.648
Radio and TV Ownership	0.736	0.711	0.876	0.792	0.771	0.751

Source: Analysis based on VLSS 1998.

Notes on targeting variables:

Poor or remote commune: 0=Commune not included in CEMMA's list of remote communes or MOLISA list of poor communes; 1=Commune included in either CEMMA difficult mountainous and remote communes or MOLISA poor communes lists;

Categories in National Poverty Map: 0= Provincial poverty headcount < 25%; 1= Headcount 25- 45%; 3= Headcount 45 -60%; 4=Headcount > 60%

Communes ranked by median expenditure: Ranking of 194 communes and urban wards in VLSS sample by median per capita expenditure of the sample households in that commune

Livestock owned: number of livestock multiplier by their livestock equivalents units: 0.7=cow, horses and water buffalo; 0.1=goats, pig and deer; 0.01=ducks and chickens

Educational Level: 0 = Post-secondary; 1=Advanced Technical; 2=Upper Secondary; 3=Lower Secondary; 4=Lower Secondary; 5=Primary; 6=Less than Primary (* Note: 1284 households do not have spouses present)

Ethnicity: 0=Kinh or Chinese Head; 1= Ethnic minority head

Floor Type: 0=Earth; 1=Other, 2=Bamboo/Wood; 3=Lime and Ash; 4=Cement; 5=Brick; 6=Marble or Tile

Roof Type: 0=Other; 1=Leaves/Straw; 2=Bamboo/Wood; 3=Canvas/Tar Paper; 4=Panels; 5=Galvanised Iron; 6=Tile; 7=Cement or Concrete

Toilet Type: 0=Flush: 1=Other: 2=None

Radio and TV Ownership: 0=Color TV; 1=Black and White TV; 2=Radio; 3 =None

---

conceal ownership of radio or televisions if it become known that their ownership would exclude household from being selected as program beneficiaries.

It would be possible to further increase the accuracy of targeting by combining a few of the above variables into a composite targeting indicator. Preliminary work on developing such an indicator using stepwise regressions shows that four variables (the number of children under 15, roof type, floor type, and the ownership of a color television), together with the choice of an appropriate poverty cut-off point, allows up to 94% of poor and non-poor households to be correctly identified in urban areas. In rural areas, developing a composite targeting indicator is more difficult, though the addition of two more variables (ethnicity and ownership of a black and white television) allows up to 75% of households to be correctly classified as poor or non-poor.[20]

6. Summary and Conclusions

Vietnam's current anti-poverty programs rely heavily on the geographic targeting of poor households. Yet, as in many developing countries, the relatively small number of households that are sampled in its national household surveys do not allow poverty statistics below the regional level to be estimated accurately. Meanwhile, questions have been raised about the comparability and reliability of the more disaggregated province, district and commune poverty statistics that are collected through Vietnam's administrative reporting system. This paper shows how the data collected by the 1998 Vietnam Living Standards Survey may be combined with that of the 1999 Population and Housing Census to bridge this gap and allow disaggregated maps of poverty to be constructed. The procedure to construct these maps involves two steps. First, the VLSS is used to explore the factors associated with poverty at the household level, and develop linear regression models for predicting per capita expenditures at the rural/urban and strata levels. Second, these regression models are applied to household data from the 3% enumeration sample of the Census to derive and map provincial level estimates of the percentage of people living in households whose per capita expenditures fall below the GSO-WB poverty line (the poverty headcount).

The national poverty map resulting from this two step procedure shows that poverty is concentrated in Vietnam's Northern Uplands, in particular in the six provinces that border China and Laos. Fourteen other provinces, most of which are located in the Northern Uplands, Central

---

[20] Further details are available from the authors on request.

Highlands and North Central Coast, have poverty headcounts above 45 percent. When rural areas are considered separately from urban areas, rural poverty is also found to be high in most of the remaining provinces of the Northern Uplands together with Gia Lai and Kon Tum and the Central Highlands. A group of moderately poor rural provinces (with rural headcounts between 45 and 50 percent) can also be seen clustered in the North Central Coast and Red River Delta. However, even relatively prosperous regions have their own pockets of poverty: such as Ha Tay in the Red River Delta and Ninh Thuan in the Southeast.

To consider the effectiveness of Vietnam's existing geographically targeted anti-poverty programs, we apply the relatively novel technique of Receiver Operating Characteristic (ROC) curves to the VLSS data. Our results confirm that a consistent ranking of communes has high potential to identify Vietnam's poor population. However, the existing officially designated list of "poor and remote communes" is less effective in targeting the poor as it excludes a large number of poor people living in other areas. Among the additional household level variables that might be used to help sharpen the focus of targeting, demographics (in particular, the number of children in a household under 15 years old), housing characteristics (especially floor type) and ownership of durable assets perform well. A simple index of radio and television ownership dominates all other individual targeting variables with the exception of communes ranked by their median per capita expenditures. Combining several household level variables into a composite targeting indicator offers the potential to further improve the targeting of the poor, especially in urban areas.

When household level data from the full sample of the 1999 Census becomes available, it should be possible to extend this poverty mapping to the district level. Since the determinants of expenditures and poverty are likely to remain relatively stable over time, we believe this will be a useful exercise even though the Census and VLSS are now three to four years out-of-date. In addition, although censuses are only conducted every ten years, the first step of the poverty mapping calculations (the expenditure regressions) can be re-estimated and new poverty maps derived, each time a nationally representative household sample survey is conducted. The complete provincial poverty map could also be redone every five years using information from the interdecadal Censuses.[21] Furthermore, international experience (Baker and Grosh, 1994, Bigman and Fofack, 2000) indicates that greater geographical disaggregation is likely to improve the targeting of Vietnam's anti-poverty programs. With more computational effort, it should also be feasible to estimate poverty headcounts (and other poverty measures too) at the commune/ward level, although it remains to be seen how accurate these calculations will be. More regionally specific analysis of the use and combination of additional household level targeting variables, such as housing characteristics and asset ownership, would also be useful at this time. Nonetheless, it is hoped that this paper has demonstrated the feasibility and policy relevance of these tools to targeting anti-poverty interventions in Vietnam.

[21] The need for updated Census data is greatest if changes in poverty are principally associated with changes in household characteristics, while the need for new households survey data is greatest if poverty changes are linked to changes in the coefficients of the expenditure regressions. Further research is needed into the relative importance of these two factors.

REFERENCES

Baker, J. and Grosh, M., 1994, “Poverty reduction through geographic targeting: how well does it work?’ *World Development*, Vol. 22, No. 7: 983-995

Baulch, B., Truong, C, Haughton, D. and Haughton, J, forthcoming, “Ethnic minority development in Vietnam: a socio-economic perspective”, Policy Research Working Papers, The World Bank, Washington D.C.

Bigman, D. and Fofack, H, 2000, *Geographic Targeting for Poverty Alleviation: Methodology and Applications*, Washington DC: World Bank Regional and Sectoral Studies

Cornia, G., and Stewart, F., 1995, “Two errors of targeting” in Van de Walle, D. and Nead. K. (eds), *Public Spending and the Poor*, Baltimore and London: John Hopkins University Press.

Elbers, C., Lanjouw, J. and Lanjouw, P., 2001, “Welfare in villages and towns: micro-level estimation of poverty and inequality”, Mimeo

Hentschel, J., Lanjouw, J., Lanjouw, P. and Poggi, J., 2000, “Combining census and survey data to trace the spatial dimensions of poverty: a case study of Ecuador”, *World Bank Economic Review*, Vol. 14, No. 1: 147-65

Minot, N., 1998, “Generating disaggregated poverty maps: An application to Viet Nam”. Markets and Structural Studies Division, Discussion Paper No. 25.. International Food Policy Research Institute, Washington, D.C.

Minot, N., 2000, “Generating disaggregated poverty maps: an application to Vietnam, 2000, *World Development*, Vol. 28, No. 2: 319-331

Minot, N. and B. Baulch, 2002, “Poverty mapping with aggregate census data: What is the loss in precision?”, Presented at the conference “Understanding Poverty and Growth in Sub-

Saharan Africa", Centre for Studies of African Economies, Oxford University, 18-20 March, 2002.

Poverty Working Group, 1999, *Vietnam: Attacking Poverty*, A Joint Report of the Government of Vietnam-Donor-NGO Poverty Working Group presented to the Consultative Group Meeting for Vietnam.

Ravallion, M. , 1992, "Poverty Comparisons", *Living Standard Measurement Working Paper* No. 88, Washington DC: World Bank.

Stata Corporation, 2001a, "Receiver operating characteristics (ROC) analysis", *Stata 7 Reference Manual,* Vol 3: 131-151, College Station, Texas: Stata Press.

Stata Corporation, 2001b, "Svymean"", *Stata 7 Reference Manual*, Vol. 4: 52-74, College Station, Texas: Stata Press.

Statistics South Africa and the World Bank, 2000, 'Is census income an adequate measure of household welfare: combining census and survey data to construct a poverty map of South Africa", Mimeo.

Van de Walle, D. and Gunewardana, 2001, "Sources of ethnic inequality in Viet Nam", *Journal of Development Economics*, Vol 65: 177-207.

Wodon, Q, 1997, "Targeting the poor using ROC curves", *World Development* Vol. 25, No 12: 2083-2092.

World Bank, 2000, *Panama Poverty Assessment: Priorities and Strategies for Poverty Reduction,* Washington DC: World Bank Country Study.

Annex 1. Descriptive statistics for variables used in regression analysis

		Rural areas				Urban areas			
Variable	Description	Mean	Std dev	Minimum	Maximum	Mean	Std dev	Minimum	Maximum
lnrpce	Log of per capita expenditure	7.56	0.478	5.879	10.148	8.293	0.602	6.526	10.732
hhsize	Size of household (members)	5.55	1.904	1.000	16.000	5.221	2.196	1.000	19.000
pelderly	Proportion over 65 yrs (fraction)	0.10	0.187	0.000	1.000	0.117	0.191	0.000	1.000
pchild	Proportion under 15 years (fraction)	0.35	0.214	0.000	0.833	0.244	0.201	0.000	0.750
pfemale	Proportion female (fraction)	0.51	0.173	0.000	1.000	0.526	0.177	0.000	1.000
ethnic	Household head is ethnic minority	0.18	0.384	0.000	1.000	0.010	0.099	0.000	1.000
Iedchd_1	Head has not completed primary school (omitted)	0.39	0.487	0.000	1.000	0.249	0.433	0.000	1.000
Iedchd_2	Head has completed primary school	0.24	0.425	0.000	1.000	0.208	0.406	0.000	1.000
Iedchd_3	Head has completed lower secondary school	0.28	0.448	0.000	1.000	0.256	0.437	0.000	1.000
Iedchd_4	Head has completed upper secondary school	0.04	0.198	0.000	1.000	0.086	0.280	0.000	1.000
Iedchd_5	Head has completed advanced technical degree	0.05	0.214	0.000	1.000	0.114	0.318	0.000	1.000
Iedchd_6	Head has post-secondary education	0.01	0.102	0.000	1.000	0.086	0.281	0.000	1.000
Iedcsp_0	Head does not have a spouse	0.14	0.344	0.000	1.000	0.207	0.405	0.000	1.000
Iedcsp_1	Spouse has not completed primary school (omitted)	0.42	0.493	0.000	1.000	0.218	0.413	0.000	1.000
Iedcsp_2	Spouse has completed primary school	0.18	0.384	0.000	1.000	0.163	0.369	0.000	1.000
Iedcsp_3	Spouse has completed lower secondary school	0.20	0.403	0.000	1.000	0.211	0.408	0.000	1.000
Iedcsp_4	Spouse has completed upper secondary school	0.03	0.173	0.000	1.000	0.056	0.229	0.000	1.000
Iedcsp_5	Spouse has completed advanced technical degree	0.03	0.163	0.000	1.000	0.090	0.287	0.000	1.000
Iedcsp_6	Spouse has post-secondary education	0.01	0.073	0.000	1.000	0.056	0.230	0.000	1.000
Ioccup_1	Head is a political leader or manager	0.02	0.126	0.000	1.000	0.032	0.176	0.000	1.000
Ioccup_2	Head is a professional or technical worker	0.03	0.163	0.000	1.000	0.100	0.300	0.000	1.000
Ioccup_3	Head is a clerk or service worker	0.05	0.212	0.000	1.000	0.264	0.441	0.000	1.000
Ioccup_4	Head is in agriculture, forestry, or fishing	0.70	0.458	0.000	1.000	0.149	0.356	0.000	1.000
Ioccup_5	Head is a skilled worker	0.07	0.259	0.000	1.000	0.190	0.392	0.000	1.000
Ioccup_6	Head is an unskilled worker	0.06	0.241	0.000	1.000	0.064	0.245	0.000	1.000
Ioccup_7	Head is not working (omitted)	0.07	0.261	0.000	1.000	0.201	0.401	0.000	1.000
Ihouse_1	House made of permanent materials	0.09	0.283	0.000	1.000	0.361	0.480	0.000	1.000
Ihouse_2	House made of semi-permanent materials	0.62	0.486	0.000	1.000	0.500	0.500	0.000	1.000
Ihouse_3	House of temporary materials (omitted)	0.29	0.456	0.000	1.000	0.139	0.346	0.000	1.000
htypla1	Interaction of log(house area) and Ihouse_1	0.34	1.108	0.000	5.537	1.417	1.914	0.000	5.835
htypla2	Interaction of log(house area) and Ihouse_2	2.35	1.876	0.000	5.293	1.832	1.865	0.000	4.973
electric	House has electricity	0.71	0.456	0.000	1.000	0.982	0.133	0.000	1.000
Inwate_1	House uses water from a public or private tap	0.02	0.136	0.000	1.000	0.578	0.494	0.000	1.000
Inwate_2	House uses well water	0.68	0.467	0.000	1.000	0.316	0.465	0.000	1.000
Inwate_3	House uses river or lake water (omitted)	0.30	0.459	0.000	1.000	0.106	0.307	0.000	1.000
Itoile_1	House has flush toilet	0.04	0.188	0.000	1.000	0.615	0.487	0.000	1.000
Itoile_2	House has latrine	0.74	0.439	0.000	1.000	0.257	0.437	0.000	1.000
Itoile_3	House has neither flush toilet nor latrine (omitted)	0.22	0.416	0.000	1.000	0.127	0.333	0.000	1.000
tv	Household has television	0.51	0.500	0.000	1.000	0.822	0.382	0.000	1.000
radio	Household has radio	0.47	0.499	0.000	1.000	0.599	0.490	0.000	1.000
reg7_1	Household in Northern Uplands (omitted)	0.20	0.403	0.000	1.000	0.092	0.290	0.000	1.000
reg7_2	Household in the Red River Delta	0.19	0.391	0.000	1.000	0.224	0.417	0.000	1.000
reg7_3	Household in the North Central Coast	0.16	0.369	0.000	1.000	0.053	0.225	0.000	1.000
reg7_4	Household in the South Central Coast	0.10	0.294	0.000	1.000	0.148	0.355	0.000	1.000
reg7_5	Household in the Central Highlands	0.05	0.212	0.000	1.000	0.000	0.000	0.000	0.000
reg7_6	Household in the Southeast	0.08	0.267	0.000	1.000	0.301	0.459	0.000	1.000
reg7_7	Household in the Mekong River Delta	0.22	0.418	0.000	1.000	0.181	0.385	0.000	1.000

Source: 1998 Viet Nam Living Standards Survey

Note: Means and standard deviations are calculated using sampling weights.

Annex 2. Determinants of per capita expenditure of each stratum

	Hanoi & HCMC		Other urban areas		Northern Uplands		Red River Delta		North Central Coast		South Central Coast		Central Highlands		Southeast		Mekong River Delta	
N	619		1111		672		783		600		502		368		514		830	
R-squared	0.4330		0.486		0.539		0.414		0.451		0.712		0.671		0.482		0.508	
Variable	coefficient	t	coefficient	t	coefficient	t	coefficient	t	coefficient	t	coefficient	t	coefficient	t	coefficient	t	coefficient	t
hhsize	-0.0688	-4.4 ***	-0.0806	-6.7 ***	-0.0835	-7.9 ***	-0.0996	-9.7 ***	-0.0758	-12.2 ***	-0.0697	-5.4 ***	-0.0587	-4.9 ***	-0.0530	-4.0 ***	-0.0876	-12.8 ***
pelderly	-0.0408	-0.4	-0.1849	-2.4 **	-0.1178	-1.3	-0.1435	-2.0 *	-0.1491	-2.4 **	-0.0006	0.0	0.0414	0.2	-0.1000	-0.5	0.0208	0.3
pchild	-0.0119	-0.1	-0.2641	-2.9 ***	-0.3242	-3.7 ***	-0.4184	-6.6 ***	-0.3163	-3.3 ***	-0.2247	-2.7 **	-0.2424	-2.2 **	-0.1399	-1.3	-0.3240	-5.3 ***
pfemale	0.0877	0.6	0.0387	0.4	-0.1101	-1.4	-0.0559	-0.7	-0.1993	-2.1 *	-0.1449	-1.4	-0.1041	-1.4	0.0521	0.6	-0.2205	-3.3 ***
ethnic	-0.2614	-1.5	0.0629	0.7	0.0220	0.4	-0.0471	-1.0	-0.0940	-1.5	-0.4229	-5.7 ***	-0.2360	-2.3 **	-0.1268	-1.0	-0.0192	-0.3
Iedchd_2	0.1198	2.1 *	0.0454	1.1	0.0394	0.8	0.0972	2.0 *	0.0152	0.4	0.0925	2.0 *	0.0092	0.2	0.0801	1.8 *	-0.0072	-0.3
Iedchd_3	0.1504	2.7 **	-0.0265	-0.8	0.0495	1.2	0.1619	3.3 ***	0.0206	0.5	0.1045	1.7	0.1235	1.7	0.0769	1.0	0.0512	1.4
Iedchd_4	0.0864	0.9	0.1437	2.7 **	0.1299	2.3 **	0.1628	3.0 ***	-0.0173	-0.3	0.0397	0.4	0.0989	0.7	0.2199	2.6 **	0.1093	1.8 *
Iedchd_5	0.1358	2.0 *	0.0725	1.2	0.0837	1.6	0.1898	3.6 ***	0.0932	1.6	0.1071	0.9	0.1929	1.0	0.2753	3.2 ***	0.1614	1.0
Iedchd_6	0.2101	2.5 **	0.1766	3.1 ***	0.1313	0.9	0.4954	4.2 ***	0.1436	2.1 *	0.4427	1.7	-0.0982	-0.3	0.0057	0.1	0.3651	2.2 **
Iedcsp_0	0.0876	1.2	-0.0087	-0.2	0.0232	0.5	0.0083	0.2	-0.0123	-0.2	-0.0034	-0.1	0.0177	0.3	0.0154	0.4	-0.0046	-0.2
Iedcsp_2	0.0996	1.3	0.0764	2.0 *	-0.0397	-1.1	-0.0149	-0.4	-0.0170	-0.4	0.0310	0.7	-0.0782	-1.1	0.0839	1.2	0.0299	0.8
Iedcsp_3	0.1423	2.0 *	0.0508	1.1	0.0219	0.6	0.0090	0.3	-0.0508	-0.9	0.1720	1.8 *	0.0401	0.5	0.1084	1.5	-0.0432	-0.7
Iedcsp_4	0.4751	3.2 ***	0.0838	1.0	0.0029	0.0	0.0203	0.4	-0.0648	-0.7	0.3033	2.3 **	-0.0851	-1.3	0.3571	2.8 **	-0.0921	-1.6
Iedcsp_5	0.1802	2.1 *	0.0091	0.2	0.1641	3.1 ***	0.0948	1.3	0.0116	0.1	-0.0565	-0.3	0.1595	1.3	-0.0083	-0.1	-0.0366	-0.1
Iedcsp_6	0.2505	3.0 ***	0.0353	0.5	0.1188	1.6	0.1520	2.8 **	0.3326	1.8 *	-0.0609	-0.3			-0.1791	-1.2	-0.1293	-0.6
Ioccup_1	0.1849	1.5	0.2371	2.7 **	0.1595	1.2	0.1464	1.7	0.0952	0.7	0.2959	5.1 ***	0.0083	0.1	0.3054	1.9 *	0.0727	0.9
Ioccup_2	-0.0377	-0.5	0.1284	1.9 *	0.1408	1.5	0.1393	1.9 *	0.0502	0.5	0.1778	1.9 *	-0.0281	-0.2	0.0619	0.5	0.2130	2.4 **
Ioccup_3	0.0192	0.3	0.0466	1.1	0.0498	0.4	0.2760	3.5 ***	0.0559	0.5	0.1103	1.1	-0.0879	-0.7	0.1426	1.3	0.0917	1.5
Ioccup_4	-0.1906	-2.8 **	0.0012	0.0	-0.0591	-0.8	0.0436	0.8	-0.0687	-1.2	0.0289	0.5	-0.0588	-0.7	-0.0829	-0.8	0.0197	0.5
Ioccup_5	-0.0614	-0.9	0.0736	1.5	0.1772	2.4 **	0.0892	1.2	0.1033	0.9	0.0028	0.0	-0.1286	-1.2	-0.0818	-0.7	0.0579	0.7
Ioccup_6	-0.1697	-1.6	-0.1292	-2.3 **	0.3607	4.7 ***	0.0204	0.2	-0.0448	-0.7	0.1239	0.6	-0.2498	-1.7	-0.2348	-2.6 **	-0.0832	-1.5
Ihouse_1	-0.8704	-4.3 ***	-0.9722	-3.5 ***	-0.0977	-0.2	-1.1440	-7.1 ***	-1.4392	-1.4	-0.4968	-0.6	0.4357	0.7	-2.7300	-2.3 **	-1.6038	-2.2 **
Ihouse_2	-0.7219	-3.6 ***	-0.5709	-3.7 ***	-0.3355	-1.6	-0.0902	-0.6	-0.2913	-1.3	-0.5064	-2.4 **	-0.1755	-0.8	-0.7977	-3.5 ***	-0.1717	-1.1
htypla1	0.2274	4.7 ***	0.3095	4.6 ***	0.0918	0.9	0.3552	8.1 ***	0.4639	1.6	0.2399	1.1	0.0545	0.4	0.6983	2.4 **	0.4052	2.6 **
htypla2	0.1850	3.6 ***	0.1826	4.6 ***	0.1233	2.2 **	0.0669	1.6	0.1126	1.7 *	0.1687	2.8 **	0.1604	3.2 ***	0.2531	4.0 ***	0.0628	1.7
electric	0.6201	3.6 ***	-0.0019	0.0	0.0217	0.3	0.1918	3.3 ***	0.0015	0.0	0.0899	1.5	0.1557	1.4	0.1725	4.4 ***	0.0903	2.4 **
Inwate_1	0.1200	1.5	0.1782	3.5 ***			0.0200	0.2	-0.4072	-5.3 ***	0.2541	2.9 **	-0.1713	-1.0	0.0189	0.1	0.1542	3.2 ***
Inwate_2	0.0073	0.1	0.0030	0.1	0.0959	1.9 *	0.1741	2.3 **	-0.1565	-2.8 **	0.0254	0.6	0.1120	4.4 ***	0.0649	2.5 **	0.1117	2.8 **
Itoile_1	0.2932	3.3 ***	0.1138	1.7 *	0.4844	6.5 ***	0.3322	3.4 ***	0.2321	1.3	0.0556	0.8	0.4115	4.1 ***	0.1856	2.7 **	0.3758	4.9 ***
Itoile_2	0.1079	0.8	-0.0152	-0.3	0.0681	1.7	0.0699	1.0	0.0444	1.3	0.0741	1.9 *	-0.0057	-0.1	0.0824	1.3	0.0567	2.0 *
tv	0.2363	3.6 ***	0.2056	4.0 ***	0.2624	10.4 ***	0.1907	8.1 ***	0.2439	5.9 ***	0.1917	4.3 ***	0.1115	2.6 **	0.2094	4.5 ***	0.1512	6.8 ***
radio	0.2558	5.5 ***	0.1573	4.6 ***	0.0313	0.9	0.0913	3.5 ***	0.1533	3.9 ***	0.0809	2.2 **	0.1415	3.1 **	0.0537	1.3	0.1492	5.2 ***
_cons	7.3886	29.8 ***	8.0018	82.1 ***	7.6097	75.2 ***	7.3747	45.1 ***	8.0240	44.2 ***	7.6878	69.5 ***	7.4845	31.5 ***	7.7554	45.2 ***	7.9655	105.8 ***

Source: Regression analysis of 1998 Viet Nam Living Standards Survey.

Note: The dependent variable is log of per capita expenditure.

* indicates that the coefficient is significant at the 10% level, ** at the 5% level, and *** at the 1% level.

Annex 3. Tests of significance of groups of explanatory variables in stratum-level regression model

Stratum	Variables	df1	df2	F statistic	Probability
Hanoi	Education of head of household	5	19	2.65	0.0557 *
and	Education of spouse	6	19	3.84	0.0112 **
HCMC	Occupation of head	6	19	6.45	0.0008 ***
	Type of housing	2	19	12.29	0.0004 ***
	Main source of water	2	19	2.24	0.1340
	Type of sanitary facility	2	19	6.09	0.0090 ***
Other	Education of head of household	5	36	3.52	0.0108 **
urban	Education of spouse	6	36	1.41	0.2364
areas	Occupation of head	6	36	3.74	0.0054 ***
	Type of housing	2	36	8.88	0.0007 ***
	Main source of water	2	36	9.24	0.0006 ***
	Type of sanitary facility	2	36	4.08	0.0252 **
Rural	Education of head of household	5	20	1.19	0.3501
Northern	Education of spouse	6	20	3.05	0.0275 **
Uplands	Occupation of head	6	20	6.13	0.0009 ***
	Type of housing	2	20	1.28	0.2986
	Main source of water	2	20	3.55	0.0743 *
	Type of sanitary facility	2	20	21.33	0.0000 ***
Rural	Education of head of household	5	24	5.99	0.0010 ***
Red	Education of spouse	6	24	1.85	0.1306
River	Occupation of head	6	24	4.54	0.0033 ***
Delta	Type of housing	2	24	25.39	0.0000 ***
	Main source of water	2	24	7.78	0.0025 ***
	Type of sanitary facility	2	24	6.06	0.0074 ***
Rural	Education of head of household	5	18	2.14	0.1071
Northern	Education of spouse	6	18	0.91	0.5103
Central	Occupation of head	6	18	3.33	0.0219 **
Coast	Type of housing	2	18	1.88	0.1811
	Main source of water	2	18	15.26	0.0001 ***
	Type of sanitary facility	2	18	1.46	0.2577
Rural	Education of head of household	5	15	1.73	0.1882
South	Education of spouse	6	15	1.69	0.1909
Central	Occupation of head	6	15	6.66	0.0014 ***
Coast	Type of housing	2	15	3.48	0.0572 *
	Main source of water	2	15	4.59	0.0278 **
	Type of sanitary facility	2	15	1.89	0.1855
Rural	Education of head of household	5	11	2.42	0.1031
Central	Education of spouse	6	11	6.79	0.0040 ***
Highlands	Occupation of head	6	11	1.23	0.3623
	Type of housing	2	11	0.67	0.5310
	Main source of water	2	11	10.68	0.0026 ***
	Type of sanitary facility	2	11	21.98	0.0001 ***
Rural	Education of head of household	5	16	3.32	0.0302 **
Southeast	Education of spouse	6	16	1.7	0.1848
	Occupation of head	6	16	5.35	0.0034 ***
	Type of housing	2	16	11.81	0.0007 ***
	Main source of water	2	16	3.07	0.0746 *
	Type of sanitary facility	2	16	3.59	0.0514 *
Rural	Education of head of household	5	25	1.95	0.1208
Mekong	Education of spouse	6	25	1.20	0.3374
River	Occupation of head	6	25	7.59	0.0001 ***
Delta	Type of housing	2	25	2.85	0.0767 *
	Main source of water	2	25	6.37	0.0058 ***
	Type of sanitary facility	2	25	12.80	0.0001 ***

Source: Regression analysis of per capita expenditure using 1998 VLSS

Note: The dependent variable is log of per capita expenditure.

* coefficient is significant at the 10% level, ** at the 5% level, and *** at the 1% level.

Annex 4: Poverty headcounts estimated with stratum-level regression

			Poverty headcount			Standard errors		
	Province code	Region	Rural	Urban	Total	Rural	Urban	Total
1	Lai Chau	NU	0.853	0.150	0.765	0.037	0.022	0.033
2	Ha Giang	NU	0.763	0.130	0.709	0.042	0.020	0.039
3	Son La	NU	0.785	0.103	0.699	0.042	0.016	0.037
4	Cao Bang	NU	0.732	0.094	0.664	0.040	0.020	0.036
5	Lao Cai	NU	0.760	0.140	0.653	0.042	0.018	0.035
6	Lang Son	NU	0.728	0.090	0.611	0.041	0.020	0.034
7	Bac Kan	NU	0.673	0.121	0.597	0.045	0.022	0.039
8	Hoa Binh	NU	0.659	0.120	0.585	0.049	0.019	0.042
9	Tuyen Quang	NU	0.638	0.115	0.581	0.051	0.016	0.045
10	Yen Bai	NU	0.661	0.114	0.554	0.047	0.017	0.038
11	Kon Tum	CH	0.689	0.217	0.533	0.062	0.029	0.043
12	Gia Lai	CH	0.642	0.193	0.532	0.060	0.025	0.046
13	Bac Giang	NU	0.561	0.121	0.530	0.060	0.018	0.055
14	Vinh Phuc	NU	0.533	0.140	0.493	0.062	0.021	0.055
15	Ninh Thuan	SE	0.561	0.208	0.480	0.045	0.028	0.035
16	Quang Tri	NCC	0.562	0.160	0.470	0.034	0.021	0.027
17	Nghe An	NCC	0.509	0.112	0.469	0.048	0.017	0.043
18	Phu Tho	NU	0.532	0.092	0.469	0.059	0.015	0.050
19	Thai Nguyen	NU	0.542	0.092	0.450	0.058	0.014	0.046
20	Ha Tinh	NCC	0.471	0.122	0.440	0.047	0.018	0.043
21	Thanh Hoa	NCC	0.467	0.112	0.435	0.043	0.017	0.039
22	Thua Thien - Hue	NCC	0.536	0.150	0.431	0.038	0.021	0.028
23	Ha Tay	RRD	0.456	0.111	0.430	0.036	0.017	0.033
24	Bac Ninh	NU	0.460	0.117	0.429	0.061	0.018	0.055
25	Quang Ngai	SCC	0.460	0.155	0.427	0.041	0.023	0.036
26	Quang Binh	NCC	0.462	0.105	0.425	0.049	0.017	0.044
27	Ninh Binh	RRD	0.453	0.095	0.408	0.036	0.015	0.031
28	An Giang	MRD	0.452	0.208	0.406	0.038	0.025	0.031
29	Ha Nam	RRD	0.421	0.120	0.403	0.035	0.018	0.033
30	Hung Yen	RRD	0.424	0.142	0.401	0.033	0.022	0.031
31	Soc Trang	MRD	0.423	0.250	0.392	0.041	0.030	0.034
32	Dong Thap	MRD	0.419	0.205	0.389	0.038	0.026	0.033
33	Dac Lac	CH	0.439	0.184	0.387	0.061	0.025	0.049
34	Tra Vinh	MRD	0.414	0.200	0.386	0.039	0.023	0.034
35	Kien Giang	MRD	0.421	0.220	0.377	0.037	0.025	0.030
36	Bac Lieu	MRD	0.422	0.221	0.374	0.036	0.027	0.028
37	Nam Dinh	RRD	0.411	0.102	0.373	0.036	0.016	0.032
38	Binh Thuan	SE	0.412	0.232	0.369	0.038	0.030	0.030
39	Phu Yen	SCC	0.410	0.192	0.368	0.042	0.027	0.034
40	Quang Nam	SCC	0.382	0.190	0.355	0.038	0.025	0.032
41	Thai Binh	RRD	0.370	0.073	0.353	0.038	0.015	0.035
42	Can Tho	MRD	0.399	0.171	0.350	0.036	0.025	0.028
43	Quang Ninh	NU	0.540	0.107	0.348	0.054	0.015	0.031
44	Ca Mau	MRD	0.387	0.167	0.347	0.035	0.023	0.029
45	Hai Duong	RRD	0.382	0.096	0.343	0.036	0.016	0.031
46	Lam Dong	SE	0.454	0.153	0.338	0.051	0.020	0.032
47	Binh Dinh	SCC	0.385	0.183	0.336	0.032	0.025	0.025
48	Vinh Long	MRD	0.360	0.162	0.332	0.035	0.024	0.030
49	Ben Tre	MRD	0.339	0.152	0.323	0.036	0.021	0.033
50	Hai Phong	RRD	0.412	0.073	0.297	0.032	0.015	0.022
51	Long An	MRD	0.321	0.166	0.296	0.038	0.021	0.032
52	Khanh Hoa	SCC	0.375	0.130	0.286	0.038	0.019	0.025
53	Tien Giang	MRD	0.283	0.123	0.263	0.037	0.018	0.033
54	Binh Phuoc	SE	0.245	0.177	0.235	0.040	0.024	0.035
55	Tay Ninh	SE	0.156	0.177	0.159	0.021	0.025	0.018
56	Da Nang	SCC	0.290	0.115	0.151	0.033	0.019	0.017
57	Ha Noi	RRD	0.331	0.014	0.149	0.032	0.005	0.014
58	Dong Nai	SE	0.155	0.122	0.145	0.022	0.020	0.017
59	Ba Ria-Vung Tau	SE	0.122	0.139	0.129	0.017	0.020	0.013
60	Binh Duong	SE	0.116	0.123	0.118	0.016	0.019	0.013
61	TP Ho Chi Minh	SE	0.096	0.038	0.048	0.022	0.012	0.011
	Total		0.441	0.110	0.365	0.010	0.002	0.011

Source Estimated from 1998 VLSS and 3% sample of 1999 Population and Housing Census

Note: A poverty headcount of 0.430 for Ha Tay implies that 43 percent of the population in Ha Tay live in households with per capita expenditures below the 1998 GSO/WB poverty line. The region codes are NU=Northern Uplands, RRD=Red River Delta, NCC=North Central Coast, SCC=South Central Coast, CH=Central Highlands, SE=Southeast, and MRD=Mekong River Delta.