MSSD DISCUSSION PAPER NO. 25

GENERATING DISAGGREGATED POVERTY MAPS: AN APPLICATION TO VIET NAM

by

Nicholas Minot

Markets and Structural Studies Division

International Food Policy Research Institute
2033 K Street N.W.
Washington, D.C. 20006
Tel. (202)-862-5600 and Fax (202)-467-4439

October 1998

Contact: Carolyn Roper
Tel. (202)-862-8148 or Fax: (202)-467-4439

This paper based on a consultancy report for the United Nations Development Programme, Capacity Building for Rural Development - Phase I. The author thanks Chris Delgado, Francesco Goletti, and Gaurav Datt for comments on various versions of the study. He also thanks Philippe Berry for assisting with data analysis and GIS processing, Nguyen Van Pham and Nguyen Dinh Chung at the General Statistics Office for processing the Agricultural Census data, and Hoang Teung Lap and Ngo The Hien at the National Institute for Agricultural Planning and Projection for assistance with the GIS processing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Geographic targeting is often recommended as a way to improve the impact of social spending and infrastructure investments on rural poverty. Previous research shows that such targeting is not very accurate unless the geographic units are small. Household surveys, however, rarely allow the estimation of poverty rates for more than 5-10 regions in a country. This study develops a method for generating disaggregated poverty maps and applies the method to Viet Nam. First, the relationship between rural poverty and 25 household indicators is estimated using household survey data. Then, census data on those same indicators are used to estimate the poverty rates for each of the 543 rural districts in Viet Nam. The results indicate that poverty is concentrated in the north and in districts furthest from the coast and cities.

1. INTRODUCTION

Over the past 15 years, researchers and policymakers have become increasingly interested in the use of targeting to improve the cost-effectiveness of programs to alleviate poverty and food insecurity. This interest has been stimulated by efforts to streamline government expenditures while protecting the poor, as well as by studies showing that a substantial share of the benefits of untargeted programs accrue to non-poor households (see Grosh, 1994).

Given the difficulty of measuring and verifying income, particularly in developing countries, one of the central issues in designing targeted programs is the trade-off between the cost of collecting information to determine eligibility and the accuracy of targeting (Glewwe and van der Gaag, 1988; Besley and Kanbur, 1993). Geographic targeting, in which eligibility is determined by the place of residence, is one of the least costly approaches. This approach is also appropriate for targeting government expenditure that has localized public-good characteristics such as investment in roads, irrigation, and marketing infrastructure.

The accuracy of regional targeting, however, varies substantially depending on the size of the geographic unit: the smaller the unit, the better the targeting (Baker and Grosh, 1994). It is not easy, however, to obtain information on poverty for a large number of small regions (e.g. districts or villages) throughout a country. Sample size constraints generally prevent the use of household surveys for estimating poverty at the neighborhood, city, or even district level. Most household budget surveys, including the World Bank Living Standards Surveys, have samples of 2000 to 6000 households, allowing estimates of poverty for just 5 to 10 regions (Grosh and Muñoz, 1996).

In Viet Nam, for example, the Ministry of Agriculture and Rural Development needs district-level information on poverty to guide the allocation of rural infrastructure projects. The 1992-93 Viet Nam Living Standards Survey, with a sample of 4800 households, provides poverty estimates for the seven agro-climatic regions, but cannot be used to estimate poverty in each of the 545 rural districts of Viet Nam.

This study develops a method to generate geographically disaggregated estimates of poverty by combining survey and census data. In the first step, I use data from the Viet Nam Living Standards Survey (VLSS) to estimate the relationship between poverty and 25 indicators (including household characteristics and regional dummy variables). In the second step, the average values of these same 25 indicators for each rural district are extracted from the 1994 Agricultural Census and are substituted into the estimated equation to generate district-level estimates of the poverty rate. The results are presented in the form of district-level poverty maps using geographic information system (GIS) software.

This paper is divided into seven sections. Section 2 reviews previous research on targeting. Section 3 describes the data and methods used in this study. Section 4 presents an overview of the patterns in various poverty indicators in Viet Nam. Section 5 examines the relationship between poverty and the poverty indicators using regression analysis. Section 6 uses this relationship to obtain estimates of the district-level poverty rates. Finally, section 7 summarizes and discusses the results of the study.

2. PREVIOUS RESEARCH

Akerlof (1978) first described the problem of designing tax and welfare policies with imperfect information about household incomes. If income cannot be observed (or can be observed only at a cost), then it may be necessary (or desirable) to set taxes and benefits according to some indicator. In the case of health and education services, the choice is between universal provision and targeted approaches. Besley (1990) explores the circumstances under which targeting is superior to universal provison of the benefit, given the administrative costs of means testing. Besley and Kanbur (1988) identify rules for the optimal distribution of benefits among categories of households assuming the government wishes to minimize one of the class of poverty measures (P_α) identified by Foster, Greer, and Thorbecke (1984). Ravallion and Chao (1989) develop an algorithm for minimizing one measure of poverty by making lump-sum transfers among a set of household categories given information about the distribution of income in each group. When applied to hypothetical regional redistribution in India and Indonesia, the impact on poverty is relatively modest, perhaps because the geographic units were fairly large (Datt and Ravallion, 1993; Ravallion, 1993).

Another issue is how to "predict" unobserved household income or expenditure based on household characteristics. Glewwe (1988) compares the targeting accuracy of various individual indicators. Glewee and Kanaan (1989), Grosh and Baker (1995), and Wodon (1997) use regression analysis to "predict" poverty as a function of various household characteristics. They find that "proxy means tests" using household characteristics can be useful for targeting.

Regional targeting is perhaps the most common type of indicator targeting, largely because it is simple and inexpensive to administer. Baker and Grosh (1994) show that state-level targeting provides only modest improvements over untargeted (uniform) transfers of the same amount of money, but district- or neighborhood-level targeting is significantly more accurate. In fact, Grosh (1994) demonstrates that programs in Latin America that focus on clinics and schools in poor neighborhoods can be as well targeted as programs that screen for eligibility based on assessments of household income. Thus, a disaggregated poverty map can allow relatively accurate targeting with minimal administrative cost. This study proposes

a method of generating disaggregated poverty maps and applies the approach to Viet Nam.

3. DATA AND METHODS

In 1992-93, the Vietnamese State Planning Committee (SPC) and the General Statistical Office (GSO) carried out the Viet Nam Living Standards Survey (VLSS) with funding from the Swedish International Development Agency and technical assistance from the World Bank. The survey used a stratified random sample of 4800 households, including 3840 rural households and 960 urban households. The 110-page questionnaire collected information on household members, housing, fertility, migration, assets, employment, agricultural production, income, and expenditure (World Bank, 1995 and SPC/GSO, 1994).

In 1994, the General Statistics Office carried out the Agricultural Census, covering 11.5 million rural households (including non-agricultural households). The five-page questionnaire collected information on household members, housing, land use, animal ownership, and assets (GSO, 1995a and GSO, 1995b). Although the Census did not attempt to collect information on income or expenditure, it does provide data on a large number of household characteristics likely to be correlated with poverty (see Table 1).

In the first step of the procedure, I use probit regression analysis with the VLSS data to estimate the probability that a rural household is poor as a function of 25 household characteristics and regional dummy variables. In probit analysis, the data are assumed to reflect the following relationship:

$$y = \Phi(\alpha + \sum_i B_i X_i + e) \qquad \textbf{(1)}$$

where y is a dependent variable taking values between 0 and 1, Φ(.) represents the cumulative density function for the standard normal curve, and the X_i are the independent variables. In this study, the dependent variable is 0 or 1 depending on whether the household is below or above the 30^{th} percentile of per capita consumption expenditure among rural households. Consumption expenditure was calculated by the author as the sum of consumption purchases, the market value of home produced food, and the rental equivalent of owner-occupied housing and major consumer durables. A regional price index, calculated as part of the VLSS study, was used to adjust for regional differences in the cost of living (World Bank, 1995). The independent variables used to "explain" poverty are household characteristics.

Table 1: Description of Poverty Indicators

Variable	Description	Available in	
		VLSS	Ag. Census
EXPPC	Value of consumption expenditure per capita (Dong/year/person)	X	
HHSIZE	Number of persons in the household	X	X
PCTAD	Percentage of household members that are adults	X	X
TAY	1 if head of household is Tay, 0 otherwise	X	
THAI	1 if head of household is Thai, 0 otherwise	X	
HOA	1 if head of household is Hoa, Chinese), 0 otherwise	X	
KHMER	1 if head of household is Khmer, 0 otherwise	X	
NUNG	1 if head of household is Nung, 0 otherwise	X	
OTHMIN	1 if head of household is other minority, 0 otherwise	X	
MINOR	1 if head of household is a minority, 0 if Kinh	X	X
MALESCH	Years of education of male adult	X	
FEMSCH	Years of education of female adult	X	
FEMHEAD	1 if head of household is female, 0 if male	X	X
FARMER	1 if main occupation is farming, 0 otherwise	X	X
FISHER	1 if main occupation is fishing, 0 otherwise	X	X
LANDPC	Annual and perennial crop land per person (sq. meters)	X	X
PCTPER	Percentage of farmland allocated to perennial crops	X	X
PCTIRR	Percentage of annual cropland that is irrigated	X	X
FOODPC	Food production per capita (kg of paddy eq. per year)	X	(1)
CATTLE	Number of cattle per household	X	X
CHICKEN	Number of chickens per household	X	X
PIGS	Number of pigs per household	X	X
HOUSAREA	Area of house (square meters)	X	X
PERM	1 if house is of permanent materials, 0 otherwise	X	X
SEMI	1 if house is of semi-permanent materials, 0 otherwise	X	X
TAP	1 if household uses water from tap, 0 otherwise	X	X
WELL	1 if household uses water from well, 0 otherwise	X	X
ELECTRIC	1 if household has electricity, 0 otherwise	X	X
RADIO	1 if household owns a radio, 0 otherwise	X	(2)
TV	1 if household owns a television, 0 otherwise	X	(2)
MBIKE	1 if household owns a motorbike, 0 otherwise	X	(2)
REG1	1 if household is in the Northern Uplands region, 0 otherwise	X	X
REG2	1 if household is in the Red River Delta region, 0 otherwise	X	X
REG3	1 if household is in the N. Central Coast, 0 otherwise	X	X
REG4	1 if household is in the S. Central Coast, 0 otherwise	X	X
REG5	1 if household is in the Central Highlands, 0 otherwise	X	X
REG6	1 if household is in the Southeast, 0 otherwise	X	X
REG7	1 if household is in the Mekong River Delta, 0 otherwise	X	X

Source: Questionnaires for Viet Nam Living Standards Survey (VLSS) and Agricultural Census.

(1) The Agricultural Census did not collect information on food production, but it has been estimated using data on foodcrop area from the Census and provincial yield estimates for 1994.

(2) The Agricultural Census refers to the number of radios, televisions, and motorbikes per 100 households. The VLSS data indicates that few than 1% of rural households have more than one radio, and the corresponding percentages for televisions and motorbikes are 0.1% and 0.3%, respectively.

The second step in the procedure is to combine the equation obtained from the regression analysis with the district-level mean values of the same poverty indicators from the Agricultural Census to generate a poverty index for each district. Because of differences in the values of the indicators between the VLSS and the Agricultural Census, a regional adjustment factor is added to ensure that the average poverty rate for each region as derived from the Census data is equal to the poverty rate in the VLSS data. Thus, regional poverty estimates are calculated as follows:

$$P_d = \Phi(\hat{\alpha} + \sum_i \hat{B}_i \bar{X}_{id} + A_r) \qquad \textbf{(2)}$$

where P_d is the district-level poverty estimate, α and B_i are the estimated parameters from the regression analysis, X_{id} is the average value of the X_i household indicator in district d, and A_r is the regional adjustment factor. Given the large number of districts, it is convenient to present the results using GIS software to produce district-level poverty maps. Sensitivity analysis is used to see whether the classification of districts varies depending on the choice of the poverty line.

Three qualifications need to be made regarding this method. First, the estimated coefficients are not necessarily consistent. Some of the indicators on the right-hand side of the regression, such as the ownership of consumer durables, are endogenous, being partly determined by per capita expenditure rather than being independent. In econometric terms, the error term (e) is likely to be correlated with some of the "independent" variables (X_i), so the estimated coefficients are subject to simultaneity bias. In addition, the probit regression model does not generally provide consistent estimates of the coefficients in the presence of heteroskedasticity. Fortunately, our focus is not on the causes of poverty or the magnitude of the coefficients but rather on developing a descriptive tool to identify poor households. In this matter, we follow the example of Glewwe and Kanaan (1989), Grosh and Baker (1995), and Wodon (1997).

The second qualification is that, because equation (2) is non-linear, the district-level poverty rate will not be exactly equal to the rate obtained by inserting district-level means into the equation, even if the equation describes household poverty without error. On the other hand, the error is likely to be small, and it is the ranking of districts and provinces, not the poverty rate itself, that is important for policy purposes.

Third, in using continuous variables for poverty indicators, this approach assumes that variation in poverty across districts will be reflected in variation in the district-level means of the indicators rather than in the distribution within districts. If differences in poverty are due mainly to differences in the distribution within districts, the continuous variables used in this study will not be very accurate. This limitation does not apply to binary (dummy) variables.

4. CHARACTERISTICS OF RURAL HOUSEHOLDS

Table 2 presents the mean value of each poverty indicator available in both the VLSS and the Agricultural Census. Regarding the poverty indicators, the two surveys yield remarkably similar results regarding the average value of household size, the proportion of adults, the percentage of farm households, the percentage of fishing households, food production, the percentage of land in perennial crops, the percentage of irrigated annual cropland, the type and size of housing, source of water, and ownership of all three consumer durables (radio, television, and motorbike).

There are some differences between the two sources: the VLSS may have overestimated land per capita and the number of livestock per household, while underestimating the proportion of households with electricity. These differences could also be due to different variable definitions or different criteria used to select households to interview.

The Agricultural Census provides information on the geographic patterns in household characteristics that are used as poverty indicators. Table 3 shows the mean values of the poverty indicators for the seven regions of Viet Nam. For example, the Northern Uplands region has the highest concentration of minority households, farmers, and small livestock. By contrast, the Red River Delta has the least land per capita, but its per capita food production (predominantly rice) is the second highest as a result of very intensive irrigation. The two central coast regions are notable for the relatively large proportion of fishing households (3.5 and 6 percent). The Central Highlands has a large minority population, abundant land, and a high proportion of land allocated to perennial crops (mainly coffee). The Southeast has a large non-agricultural population and the highest rates of ownership of the three consumer durables, reflecting its urbanized high-income population. Finally, the Mekong River Delta is notable for being relatively land abundant and highly irrigated, contributing to its per capita food production which is substantially higher than that of any other region.

More disaggregated geographic patterns can be obtained from the Agricultural Census. For example, Figures 1-3 show the district-level means of three of the poverty indicators used in this study. Figure 1 reveals that the proportion of ethnic minority households is highest in the north and in the Central Highlands. It is notable that the line between minority districts and non-minority districts is quite sharp: in most districts the proportion of ethnic minorities is either less than 5 percent or more than 50 percent. Figure 2 shows the level of food production per capita, expressed in paddy equivalent. Not surprisingly, food production is highest in the two deltas but the pattern is somewhat erratic with pockets of intensive food production elsewhere. Figure 3 shows that television ownership is concentrated around Hanoi and in the south.

Table 2: Average Value of Poverty Indicators in Rural Areas

Indicator	Viet Nam Living Standards Survey	Agricultural Census
Expenditure per capita (1000 D/yr)	100.4	NA
Number of persons in household	5.0	4.8
Percentage of adults 16-60 yrs	51.8	52.6
Percentage of female heads	22.6	22.3
Percentage of minority heads	14.2	12.4
Education of male adults (years)	6.7	NA
Education of female adults (years)	4.9	NA
Percentage of farming households	78.5	79.5
Percentage of fishing households	2.4	1.9
Agricultural land per capita (m^2)	1009.3	870.7
Percentage of land in perennial crops	13.2	13.6
Percentage of households without land	7.8	14.5
Percentage of annual cropland irrigated	50.4	46.6
Food production (kg paddy eq./person/yr)	418.0	442.6 (1)
Number of cattle per household	0.6	0.3
Number of chickens per household	13.2	7.5
Number of pigs per household	1.6	1.3
Area of house (m^2)	41.6	39.3
Percentage in permanent houses	10.8	11.8
Percentage in semi-permanent houses	48.4	43.3
Percentage with water from tap	1.2	1.1
Percentage with water from well	62.4	64.8
Percentage with electricity	38.8	52.4
Percentage with a radio	33.8	37.0 (2)
Percentage with a television	15.3	21.0 (2)
Percentage with a motorbike	6.0	8.5 (2)
Percentage living in Northern Uplands	17.5	17.3
Percentage living in Red River Delta	25.0	23.4
Percentage living in N.C. Coast	15.0	15.3
Percentage living in S.C. Coast	10.0	10.7
Percentage living in C. Highlands	3.3	4.2
Percentage living in Southeast	8.3	8.2
Percentage living in Mekong Delta	20.8	21.0

Source: Viet Nam Living Standards Survey and Agricultural Census.
Note: (1) See note 1 in Table 1.
(2) See note 2 in Table 1.

Table 3: Poverty Indicators in Rural Areas by Region

Indicator	Northern Uplands	Red River Delta	N. Central Coast	S. Central Coast	Central Highlands	Southeast	Mekong River Delta
HHSIZE (nbr.)	5.1	4.1	4.7	4.8	5.2	5.0	5.1
PCTAD (% hh members)	50.3	53.3	49.2	52.8	50.2	56.2	55.3
FEMHEAD (% hh)	17.0	23.6	19.6	26.8	17.7	28.2	23.2
MINOR (% hh)	38.9	0.3	9.0	6.3	36.6	5.5	7.8
FARMER (% hh)	91.2	91.1	83.3	74.4	76.8	49.0	69.9
FISHER (% hh)	0.3	0.3	3.5	6.0	0.1	1.7	2.3
LANDPC (m^2)	787.8	519.8	550.7	661.7	1232.0	996.6	1438.8
PERPCT (% land)	6.0	2.2	4.2	6.6	35.2	41.9	13.8
FOODPC (kg/yr)	255.3	383.7	251.2	313.8	187.2	209.9	772.7
IRRIG (% land)	26.6	82.5	51.5	36.8	8.6	19.0	54.2
CATTLE (nbr/hh)	0.3	0.1	0.4	0.7	0.7	0.2	0.1
CHICK (nbr/hh)	12.4	7.9	9.0	4.8	5.0	6.1	4.3
PIGS (nbr/hh)	1.8	1.3	1.4	1.2	1.2	0.8	0.8
HOUSAREA (m^2)	44.6	33.2	39.1	37.7	33.1	42.8	42.8
PERM (% hh)	9.6	24.5	10.4	7.2	4.6	3.4	7.6
SEMI (% hh)	51.9	49.4	56.3	45.0	48.2	43.2	18.7
TAP (% hh)	0.3	1.0	0.4	0.9	0.3	4.7	1.2
WELL (% hh)	70.6	67.1	85.5	89.8	72.8	84.7	22.3
ELECTRIC (% hh)	50.3	89.0	55.5	46.3	18.8	44.1	24.2
RADIO (nbr/100 hh)	43.9	35.7	34.3	31.3	31.6	44.8	35.8
TV (nbr/100 hh)	20.1	23.4	11.0	14.8	14.7	33.3	26.3
MBIKE (nbr/100 hh)	6.7	5.9	3.7	11.1	13.1	27.0	7.4

Source: 1994 Agricultural Census.

Table 4: Poverty Indicators for Rural Households by Expenditure Category

Variable	Poorest 25%	2nd quartile	3rd quartile	Richest 25%	All rural households
REXP_PER (d/mon)	46.9	70.8	98.3	185.6	100.4
HHSIZE (nbr.)	5.5	5.1	4.8	4.3	4.9
PCTAD (% hh members)	45.8	50.4	54.0	56.7	51.7
MINOR (% hh)	26.4	14.6	9.9	5.7	14.1
FEMHEAD (%)	22.4	20.1	20.9	27.0	22.6
MALESCH (yrs)	5.8	6.7	7.1	7.4	6.7
FEMSCH (yrs)	4.3	5.1	4.4	5.1	4.8
FARMER (% hh)	84.9	84.3	76.2	68.4	78.4
FISHER (%hh)	1.8	2.3	3.0	2.4	2.4
LANDPC (m^2)	734.2	845.8	1038.1	1418.9	1009.2
PCTPER (% land)	11.6	13.0	13.0	15.0	13.1
FOODPC (kg/yr)	263.8	359.2	461.8	586.9	417.9
CHICKEN (nbr/hh)	10.1	14.8	14.3	13.2	13.1
CATTLE (nbr/hh)	0.6	0.5	0.5	0.4	0.5
PIGS (nbr/hh)	1.3	1.6	1.6	1.7	1.6
HOUSAREA (m^2)	36.1	40.3	42.8	46.8	41.5
SEMI (% hh)	38.0	49.1	51.3	54.9	48.3
PERM (% hh)	5.1	10.3	12.8	15.1	10.8
TAP (% hh)	0.1	0.5	1.4	2.9	1.2
ELECTRIC (% hh)	25.3	38.4	42.9	48.4	38.7
TV (% hh)	4.1	10.0	18.7	28.2	15.2
RADIO (% hh)	21.3	30.0	38.9	44.9	33.8
MBIKE (% hh)	1.3	3.7	5.0	14.0	6.0
Northern Uplands (%)	24.0	20.2	14.8	10.1	17.5
Red River Delta (%)	18.4	27.0	30.6	23.4	25.0
N. Central Coast (%)	20.7	17.3	12.2	9.8	15.0
S. Central Coast (%)	10.6	7.0	9.9	12.5	10.0
Central Highlands (%)	4.0	3.3	2.4	3.6	3.3
Southeast (%)	5.9	6.6	8.4	12.4	8.3
Mekong River Delta (%)	15.4	18.1	21.7	28.1	20.8

Source: Viet Nam Living Standards Survey.

Figure 1: Ethnic minority households (percentage)

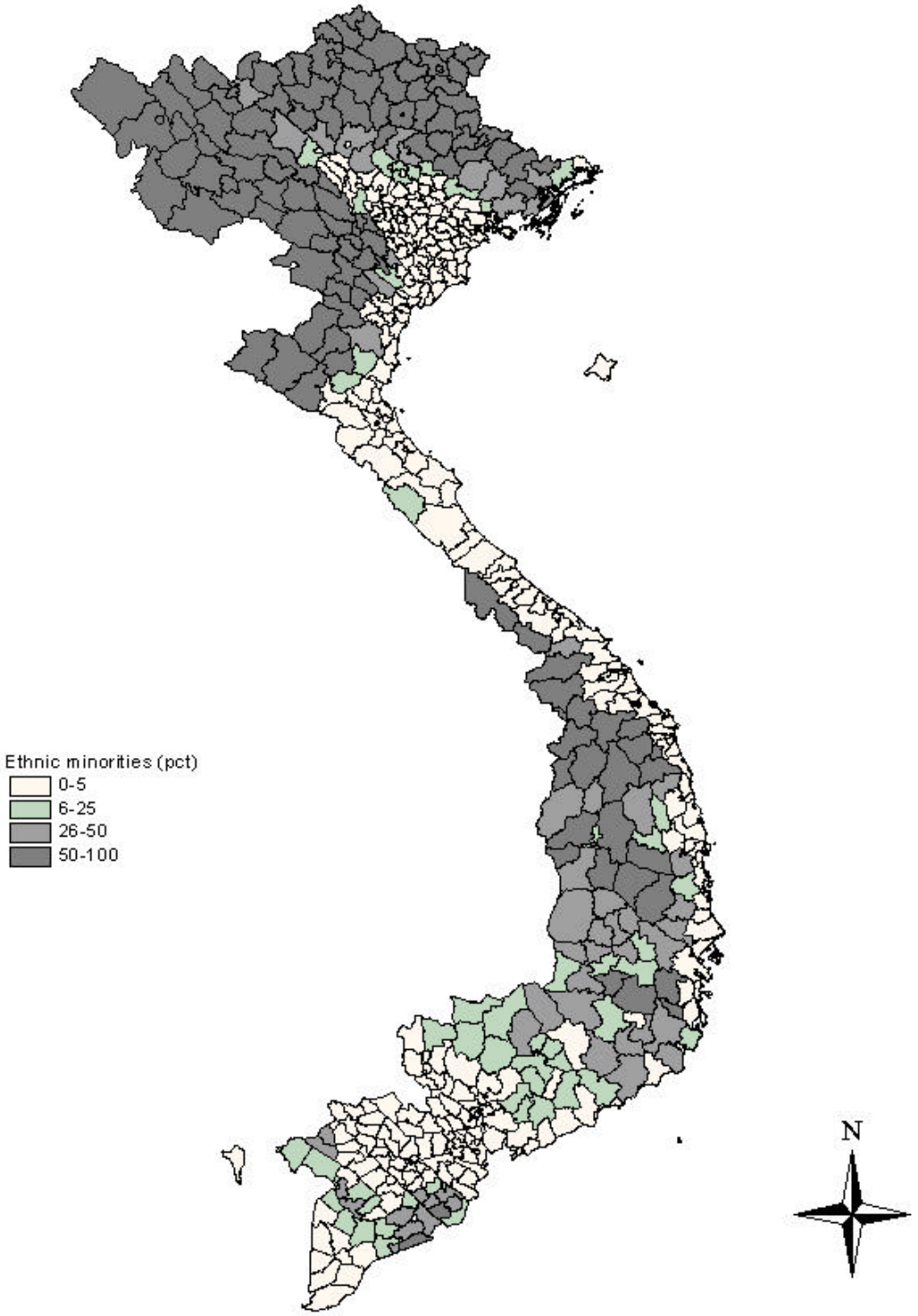


Figure 2: Food production in paddy equivalent (kg/person/year)

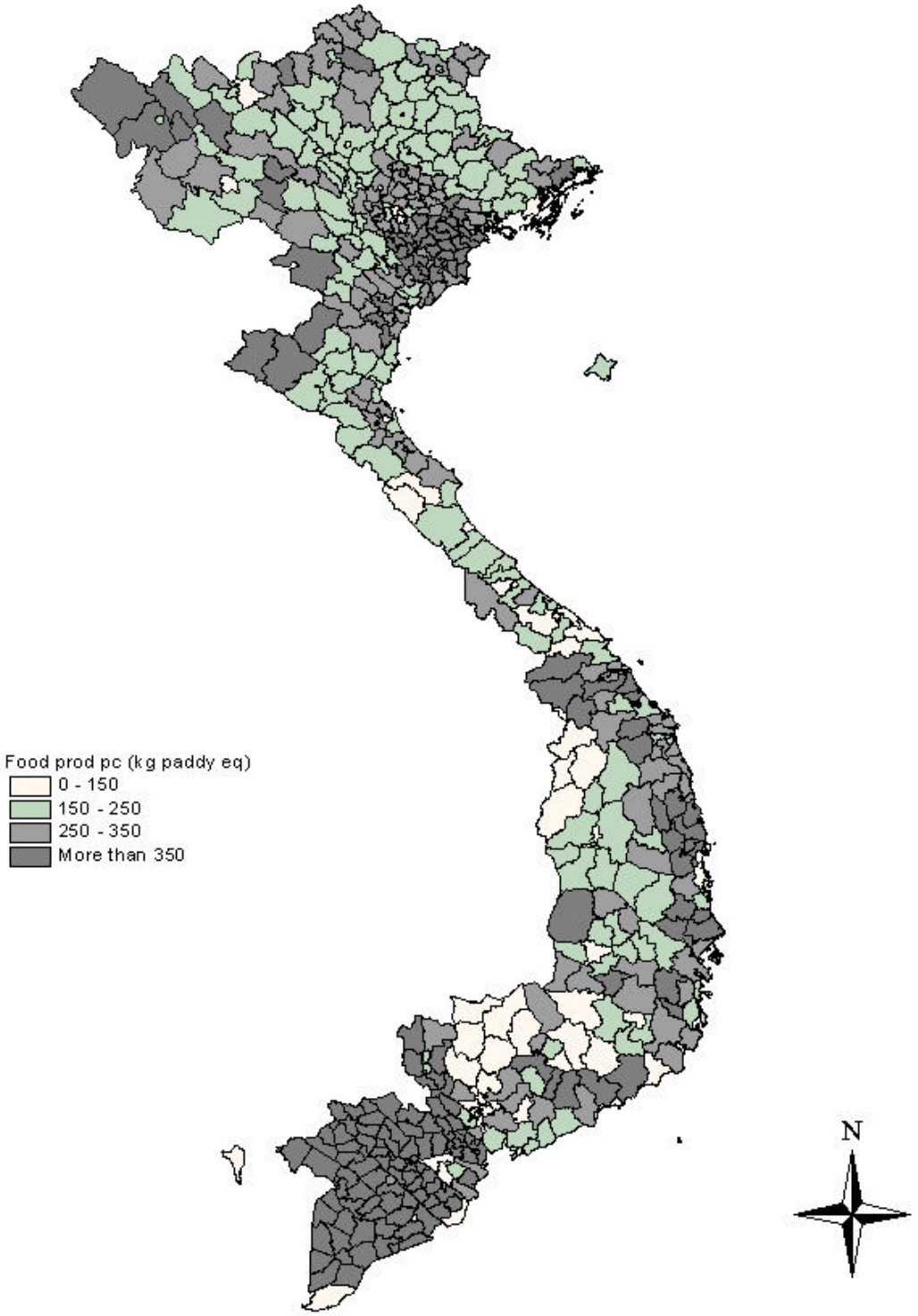
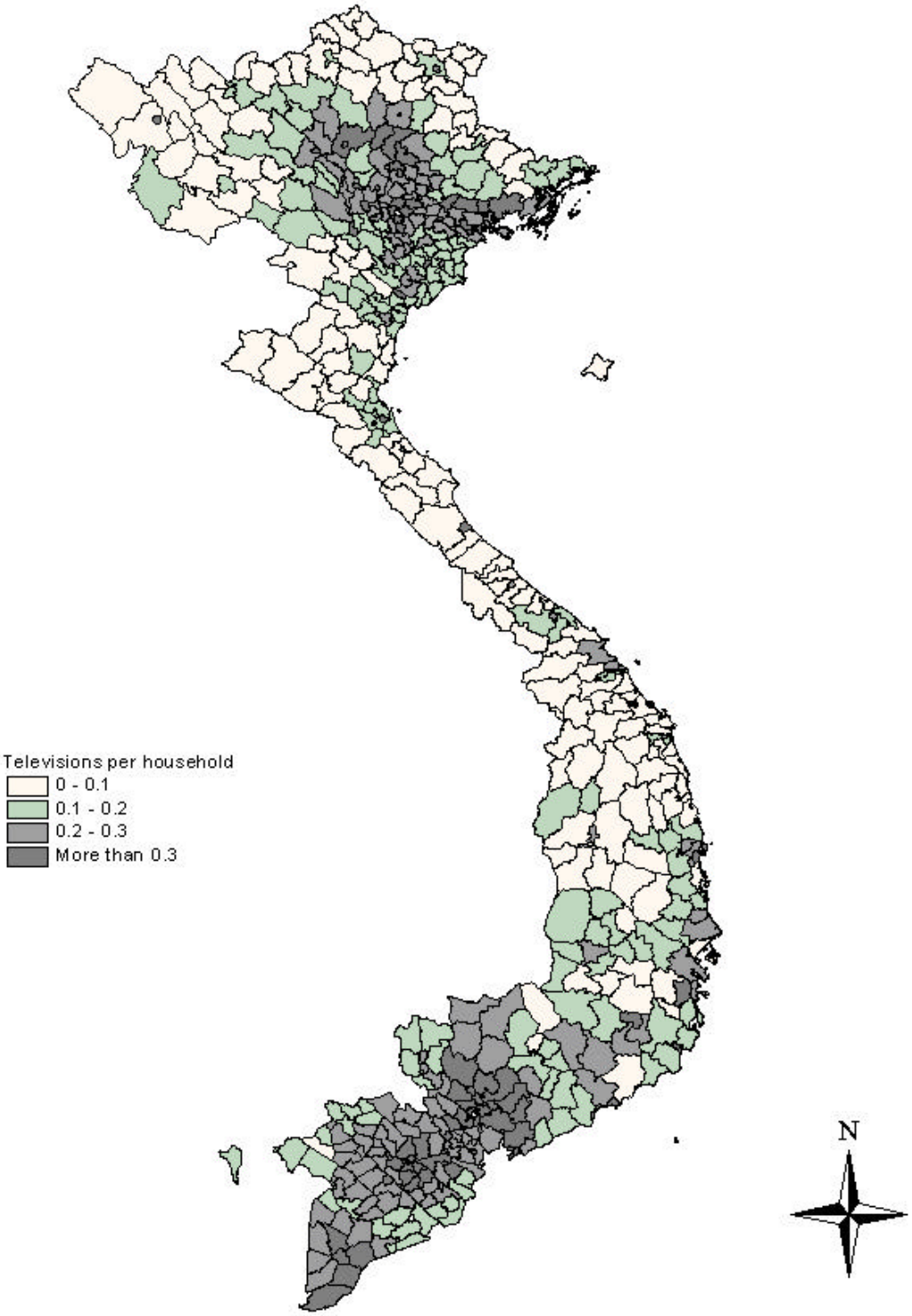


Figure 3: Television ownership (sets per household)



The advantage of the VLSS data is that it allows us to examine the relationships between poverty and household characteristics. Table 4 shows the average value of the poverty indicators for rural households in different expenditure quartiles, based on the VLSS data. The table shows that households in the higher expenditure categories tend to have smaller households and a larger share of adults among household members. The proportion of households headed by a minority falls from 26 percent in the poorest category to less than 6 percent in the richest category. In contrast, there is no consistent pattern with regard to female-headed households. The table also shows that poor households are characterized by lower levels of schooling among adult males. It is harder to identify a clear pattern in female schooling across expenditure categories.

While the bulk of rural households have farming as their main occupation, the proportion is highest among the two poorer categories. Presumably, the two richer categories have more households earning income from non-farm self-employment and wages.

Not surprisingly, the amount of land per capita, food production per capita, and the number of pigs per household tend to rise across expenditure categories. Higher income households also devote a larger share of their crop land to perennial crops (primarily coffee and fruit trees). The number of chickens per household increases between the poorest category and the second category, but does not increase across the higher income categories. This suggests that the number of chickens may be more useful for identifying the poorest 25 percent than for identifying the poorest 50 percent of rural households.

House size, electrification, access to tap water are all positively associated with expenditure per capita. Similarly, the ownership of consumer goods (radios, televisions, and motorbikes) is strongly correlated with expenditure per capita. Thus, Table 4 provides tentative support for the idea that these variables may serve as poverty indicators. In the next section, regression analysis is used to study the strength of these relationships, while controlling for the value of other variables.

5. RELATIONSHIP BETWEEN POVERTY AND INDICATORS

What is the relationship between poverty and various poverty indicators? First, we examine the accuracy of using each indicator separately. Then a probit analysis is used to "predict" poverty as a function of all the indicators, based on the 3840 rural households in the 1992-93 Viet Nam Living Standards Survey (VLSS).

Targeting with Individual Indicators

Table 5 shows the accuracy of 22 poverty indicators taken individually. The targeting criterion, in the second column, is the rule used to predict which households are poor. For continuous variables, the criterion is set so that the proportion of household that qualify is 30 percent, equal to the defined poverty rate. "Leakage," in the third column, is the proportion of VLSS rural households predicted to be poor by the targeting rule that are in fact not poor. "Undercoverage" refers to the proportion of poor households that are not identified as poor by the targeting rule. An untargeted program (providing benefits to all households) would have a leakage rate of 70 percent (given our defined poverty rate of 30 percent) and an undercoverage rate of zero percent. Perfect targeting would imply leakage and undercoverage rates of zero.

Table 5 shows that none of the indicators is, by itself, very successful at identifying poor households. The leakage rate is generally between 60 and 70 percent, meaning that 60-70 percent of the beneficiaries under these targeting rules would not be poor. The best leakage rate is achieved by targeting ethnic minority households since only 48 percent of them are not poor (in other words, 52 percent of the rural ethnic minority households fall below the 30^{th} percentile of rural households). Only two other indicators (region and proportion of working-age adults in the household) also achieve leakage rates below 60 percent.

The rates of undercoverage vary widely, from less than 5 percent to over 75 percent. The lowest rates are achieved by using targeting rules that cover the vast majority of rural households (e.g. household that do not have televison and households that do not have a motorbike). These same rules also have leakage rates close to 70 percent, the rate that would result from providing benefits to all households (no targeting).

The most common poverty indicator used in Viet Nam is food production per capita, expressed in paddy equivalent (Anh, 1997 and Huu, 1997). Although this may have been a useful indicator in the 1980s when food shortages were common, Table 5 shows that per capita food production is not a particularly good indicator, with leakage and undercoverage rates of 63.8 percent each. The only indicator that has leakage and undercoverage rates below 60 percent is region, where the criterion is to target households living in the Northern Uplands and the North Central Coast.

Table 5: Targeting Accuracy of Various Poverty Indicators

Variable	Targeting criterion	Leakage (%)	Undercoverage (%)
HHSIZE	Household has 6 or more members	61.2	53.6
PCTAD	Less than 40% of the members are of working age	59.0	67.0
MINOR	Household head is ethnic minority	47.7	75.3
MALESCH	Head male has 4 years or less of education	61.8	67.1
FEMSCH	Head female has 2 years of less of education	63.5	60.9
FEMHEAD	Household head is female	69.5	77.0
FARMER	Main occupation of household is farming	67.7	15.5
FISHER	Main occupation of household is not fishing	69.8	1.9
LANDPC	Household has less than 477 ares/person of farmland	63.8	63.8
PCTPER	Household has no perennial crops	67.3	49.0
FOODPC	Household produces less than 175 kg/person of food	63.9	63.9
CATTLE	Household has no cattle	72.0	39.3
CHICKEN	Household has less than 2 chickens	67.2	69.4
PIGS	Household has no pigs	66.4	67.1
HOUSAREA	Area of the house is less than 29 m^2	62.9	62.8
PERM	Household does not have permanent housing	68.4	6.0
SEMI	Household does not have semi-permanent housing	65.1	39.8
TAP	Household does not get water from a tap	69.7	0.3
TV	Household does not have a television	66.3	4.9
RADIO	Household does not have a radio	64.8	22.3
MBIKE	Household does not have a motorbike	68.6	1.6
REGION	Household lives in N. Uplands or N. Central Coast	58.7	55.3

Source: Calculated from the Viet Nam Living Standards Survey.

Notes: Leakage refers to the proportion of beneficiaries (households meeting the targeting criterion) that are not poor (households below the 30^{th} percentile). Undercoverage refers to the proportion of poor households that are not beneficiaries "Food" is the amount of paddy, maize, cassava, and sweet potatoes expressed in paddy equivalent.

Targeting with Combined Indicators

A probit regression model is used to "predict" household poverty as a function of a group of indicators. Model 1 includes the full range of potential poverty indicators in the first column of Table 1. Model 2 uses a subset of indicators that were useful in helping to “predict” whether or not a household is poor. Model 2 is also limited to indicators that are available from the Agricultural Census, since only variables that are in both the VLSS and the Agricultural Census can be used to generate the district-level poverty map.

The improvement in targeting accuracy can be shown by calculating the leakage and undercoverage rates for a program that targets the poorest 30 percent of households as determined by their *predicted* poverty rate according to these two models. With Model 1, leakage and undercoverage rates are just 17.3 percent. With Model 2, the two rates are 17.6 percent. In other words, under either model, less than 18 percent of those identified as poor by the model are actually not poor (leakage) and less than 18 percent of those that are poor are not identified as such by the model.

Tables 6 and 7 provide the results of Models 1 and 2, respectively. The first three columns show the coefficients, t-statistics, and the estimated probability that the true coefficient is zero for each coefficient. The fourth column of each table shows the marginal effect of each indicator on the probability that a household is poor, evaluated at the means of all variables.

In both models, large households and households with a small proportion of adults are more likely to be poor. For example, the fourth column indicates that increasing the household size by one is associated with an 5 percentage point increase in the probability that a household is poor.

Being an ethnic minority is associated with a 17 percentage point increase in the probability that a household is poor. These ethnic differences presumably reflect language barriers, physical isolation, quality of agricultural land, and other missing variables.

Model 1 shows that households in which the male adult has more education are less likely the household is to be poor. On the other hand, the association between poverty and female education is not statistically significant in this analysis. It should be recognized, however, that female education may effect household welfare in ways not picked up by consumption expenditure, such as through better nutrition or health. Model 2 excludes education because this indicator is not available in the Agricultural Census data.

Table 6: Probit Regression Analysis of Poverty (Model 1 with All Indicators)

Variable	Coefficient	Coefficient/ standard error	Probability that coefficient = 0	Marginal effect
Constant	0.1883	1.314	0.1887	0.0621
HHSIZE	-0.1543	-11.537	0.0000	-0.0509
PCTAD	0.7049	6.169	0.0000	0.2326
TAY	0.0005	0.000	0.9973	0.0002
THAI	-0.5656	-2.461	0.0139	-0.1867
HOA	0.5371	1.055	0.2914	0.1773
KHMER	-0.6535	-3.697	0.0002	-0.2157
NUNG	-0.5190	-2.745	0.0061	-0.1713
OTHMIN	-0.6133	-4.636	0.0000	-0.2024
MALESCH	0.0496	-5.671	0.0000	0.0164
FEMSCH	-0.0340	-0.425	0.6711	-0.0112
FEMHEAD	-0.1967	-3.181	0.0015	-0.0649
FARMER	-0.3532	-5.068	0.0000	-0.1166
FISHER	0.5512	3.347	0.0008	0.1819
LANDPC	0.0002	5.349	0.0000	0.0001
PCTPER	0.2103	2.088	0.0368	0.0694
FOODPC	0.0006	5.922	0.0000	0.0002
CATTLE	0.0480	2.178	0.0294	0.0158
CHICKEN	0.0029	2.186	0.0288	0.0010
PIGS	0.0219	1.882	0.0598	0.0072
HOUSAREA	0.0069	4.539	0.0000	0.0023
PERM	0.6479	6.132	0.0000	0.2138
SEMI	0.3252	5.348	0.0000	0.1073
TAP	0.9495	2.931	0.0034	0.3134
WELL	0.0833	1.389	0.1648	0.0275
ELECTRIC	0.0854	1.373	0.1699	0.0282
RADIO	0.3218	5.660	0.0000	0.1062
TV	0.3880	4.260	0.0000	0.1280
MBIKE	0.4059	2.796	0.0052	0.1340
REG1	-0.8804	-8.138	0.0000	-0.2905
REG2	-0.4677	-4.595	0.0000	-0.1544
REG3	-0.7296	-6.851	0.0000	-0.2408
REG4	-0.2242	-1.910	0.0561	-0.0740
REG5	0.0978	0.550	0.5824	0.0323
REG6	-0.3105	-2.677	0.0074	-0.1025
Log likelihood ratio		0.222		
Significance level		0.000		

Source: Data from Viet Nam Living Standards Survey. Regression carried out with LIMDEP and SPSS software.

Note: The number of observations is 3840.
The dependent variable is 0 if the household is among the poorest 30 percent; 1 if otherwise.
The variables are defined in Table 1.

Table 7: Probit Regression Analysis of Poverty (Model 2 with Selected Indicators)

Variable	Coefficient	Coefficient/ standard error	Probability that coefficient = 0	Marginal effect
Constant	0.4428	3.3370	0.0008	0.1461
HHSIZE	-0.1535	-11.6310	0.0000	-0.0506
PCTAD	0.7311	6.4200	0.0000	0.2413
MINOR	-0.5277	-6.7230	0.0000	-0.1742
FEMHEAD	-0.1368	-2.3040	0.0212	-0.0451
FARMER	-0.4140	-6.0600	0.0000	-0.1366
FISHER	0.4316	2.6610	0.0077	0.1424
LANDPC	0.0002	4.8850	0.0000	0.0001
PCTPER	0.3188	3.2100	0.0013	0.1052
FOODPC	0.0007	6.3230	0.0000	0.0002
CATTLE	0.0416	1.9240	0.0543	0.0137
CHICKEN	0.0032	2.4040	0.0162	0.0010
PIGS	0.0250	2.2060	0.0273	0.0083
HOUSAREA	0.0064	4.3300	0.0000	0.0021
PERM	0.6652	6.6090	0.0000	0.2195
SEMI	0.3739	6.3730	0.0000	0.1234
TAP	0.8989	2.7870	0.0053	0.2966
TV	0.4271	4.8050	0.0000	0.1409
RADIO	0.3492	6.2250	0.0000	0.1152
MBIKE	0.4583	3.2110	0.0013	0.1512
REG1	-0.6039	-6.4500	0.0000	-0.1993
REG2	-0.2517	-2.8570	0.0042	-0.0831
REG3	-0.5354	-5.7840	0.0000	-0.1767
REG4	-0.1313	-1.2680	0.2048	-0.0433
REG5	0.1009	0.6640	0.5069	0.0333
REG6	-0.2218	-2.0350	0.0418	-0.0732
Log likelihood ratio		0.210		
Significance level		0.000		

Source: Data from Viet Nam Living Standards Survey. Regression carried out with LIMDEP and SPSS software.

Note: The number of observations is 3840.
The dependent variable is 0 if the household is among the poorest 30 percent; 1 if otherwise.
The variables are defined in Table 1.

Holding other factors constant, female-headed households are about 5 percentage points more likely to be poor than are male-headed households. Farming households are more likely to be poor, whereas fishing households are less likely to be poor.

Agricultural production patterns are also correlated with poverty. Poverty is associated with having less land per capita, lower food production per capita, and a smaller share of land allocated to perennial crops. Similarly, households with small numbers of chickens, pigs, and cattle are more likely to be poor.

Housing characteristics also show a statistically significant relationship with the likelihood that a household is poor. Families living in permanent houses are less likely to be poor than those in semi-permanent houses, while the latter are less likely to be poor than those in temporary houses. House size is also negatively correlated with poverty. The very large coefficient of the variable indicating tap water suggests that it is very unlikely that a household with tap water is poor.

Ownership of a the three consumer durables are also significant indicators of poverty. One would expect these indicators to be particularly useful because they are less affected by occupation, farm type, and region than are the agricultural indicators such as animal ownership, food production, and perennial crop area.

Finally, the regional dummy variables in the probit model indicate that poverty rates vary across the seven agro-climatic regions even after controlling for the above factors. In particular, the north and central coast of Viet Nam are poorer than Mekong River Delta, even after controlling for the other variables.

A few variables did not have a statistically significant relationship to poverty after controlling for the other indicators: irrigation, access to electricity, and access to well water. The effect of irrigation is probably already captured by the variable representing per capita food production. Similarly, electricity only becomes statistically significant when the television variable is excluded from the analysis. And the difference between households using well water and those using river or lake water is small, so access to well water is not a good indicator of poverty.

6. ESTIMATING DISTRICT-LEVEL POVERTY RATES

In this section, I calculate district-level poverty by combining the regression results from Model 2 and data from the 1994 Agricultural Census. The regression analysis gives us an equation that describes the probability a household is poor as a function of 19 household characteristics and 6 regional dummy variables. I insert district-level mean values of the same poverty indicators into this equation to get an estimate of the poverty rate of the district. The districts are then ranked to identify four poverty categories and mapped using GIS software.

Figure 4 shows the resulting district-level poverty map for Viet Nam. The lightest shade identifies the poorest 50 districts. The map reveals that these 50 districts are concentrated in the northern and western edges of the Northern Uplands region, the western edge of the North Central Coast, and in the northern part of the Central Highlands. In general, these are hilly areas that are far from the large cities and far from the coast. They are also areas with large ethnic minority populations.

The darkest shade represents the least poor rural districts. In the north, these districts correspond almost exactly to the Red River Delta region, although they also include parts of the Northern Uplands region. In the Central Highlands, 10 districts in Dak Lak and 8 districts in Lam Dong fall into this category. In the Southeast, most of the districts are relatively prosperous. Similarly, in the Mekong River Delta most of the districts are relatively well off except for a few coastal districts and one on the Cambodian border.

Table 8 shows the distribution of districts according to region and poverty category. Of the poorest 50 districts, 32 (64%) are in the Northern Uplands and 12 (24%) are in the North Central Coast. There are also four of these districts in the Central Highlands and two in the South Central Coast region.

Table 9 shows the rank of provinces by estimated poverty level.[1] According to these calculations, Lai Chau and Lao Cai on the northern border are the two poorest provinces. Every district in these two provinces falls among the poorest 150 districts in the country. The eight poorest provinces are all located in the Northern Uplands. All the districts in these provinces are among the 250 poorest in Viet Nam. Also among the ten poorest provinces are Kon Tum and Gia Lai, both in the Central Highlands.

The regression analysis that generated our poverty map uses a poverty line set at the 30th percentile of rural households. In other words, our definition of poverty assumes that 30 percent of rural households are poor. This is a reasonable but arbitrary poverty line, so it is worth asking whether the results are sensitive to alternative poverty lines. The regression analysis was repeated using poverty lines at the 25th and 35th percentile of consumption expenditure. Some variables (such as electricity) become statistically significant while others (such as cattle) lose their statistical significance, but the results and the overall explanatory power of the different versions are similar. District-level poverty estimates were calculated using these alternative poverty lines and the classification compared. As shown in Table 10, using poverty lines at the 25th and 35th percentile had little effect on the

[1] The estimated poverty rate for each province is calculated as a weighted average of the poverty rates for the districts in the province, where the weights are the proportion of households in each district.

classification of districts into poverty groups. Using a poverty line at the 25th percentile gives the same classification for 489 (90 percent) of the 543 rural districts. Using a poverty line at the 35th percentile gives the same classification for 526 (97 percent) of the 543 districts.

Table 8: Distribution of Districts by Region and Poverty Category (Number of Districts)

Region	Poorest 50 districts	2nd poorest group (100 districts)	3rd poorest group (100 districts)	Least poor group (293 districts)	Total
Northern Uplands	32	61	26	24	143
Red River Delta	0	0	2	69	71
N. Central Coast	12	14	36	12	74
S. Central Coast	2	10	17	36	65
Central Highlands	4	13	7	21	45
Southeast	0	0	4	39	43
Mekong River Delta	0	2	8	92	102
Total	50	100	100	293	543

Source: Analysis of data from 1994 Agricultural Census (see text).
Note: There were 545 rural districts in the 1994 Agricultural Census, but because of missing data this analysis estimates poverty for just 543 rural districts.

Table 9: Provinces with the Highest Poverty Rates

Rank	Province	Region	Distribution of districts by poverty category				
			Poorest 50 districts	2nd poorest group (100 districts)	3rd poorest group (100 districts)	Least poor group (293 districts)	Total
1	Lai Chau	Northern Uplands	5	4	0	0	9
2	Lao Cai	Northern Uplands	6	4	0	0	10
3	Son La	Northern Uplands	6	3	1	0	10
4	Ha Giang	Northern Uplands	5	4	1	0	10
5	Cao Bang	Northern Uplands	5	8	0	0	13
6	Hoa Binh	Northern Uplands	1	9	0	0	10
7	Yen Bai	Northern Uplands	1	6	1	0	8
8	Lang Son	Northern Uplands	2	8	1	0	11
9	Kon Tum	Central Highlands	3	3	0	1	7
10	Gia Lai	Central Highlands	1	6	2	2	11

Source: Calculated from 1994 Agricultural Census (see text).

Figure 4: Poverty (estimated from household characteristics)

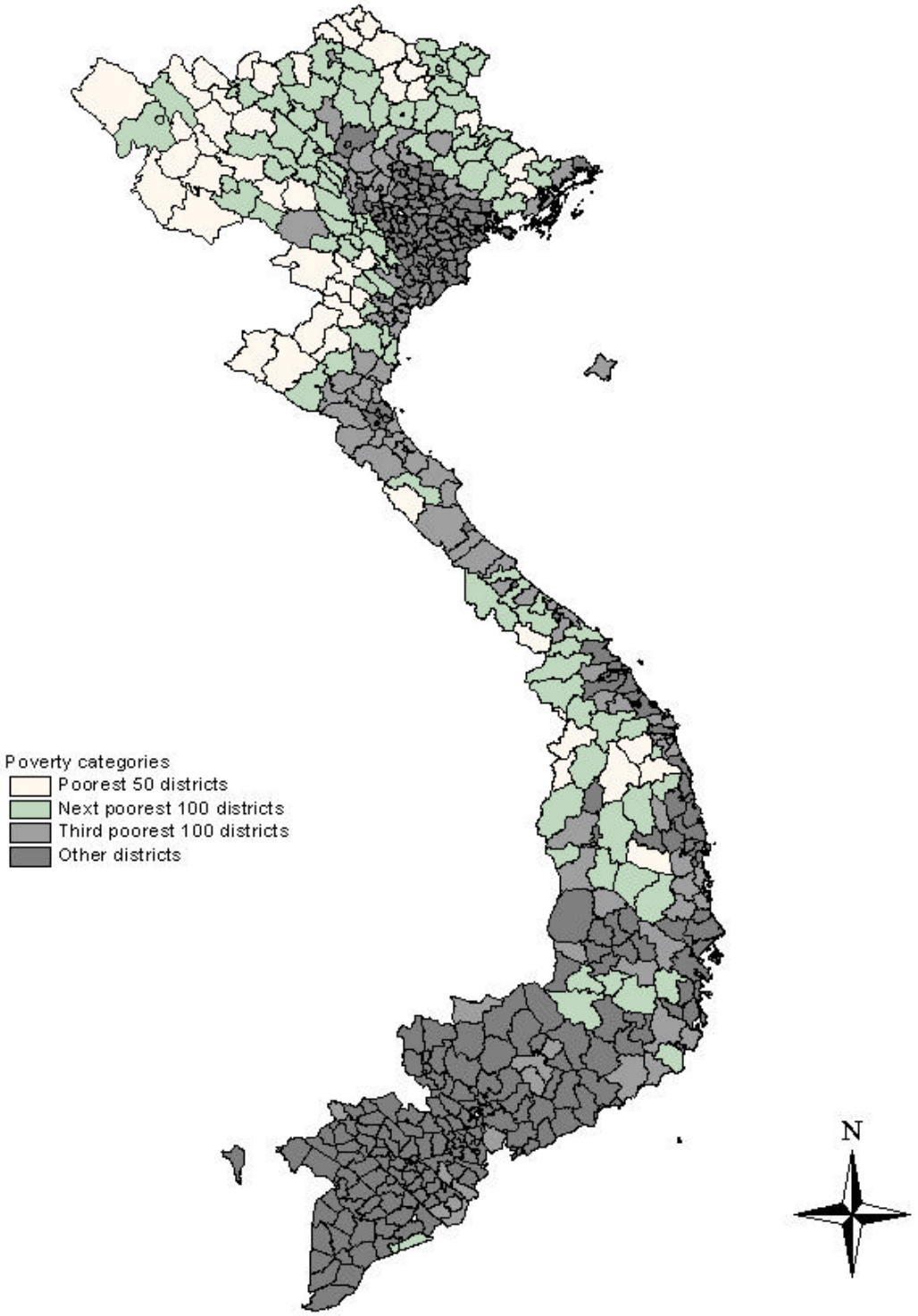


Table 10: Sensitivity of the Poverty Classification of Districts to the Poverty Line

Classification of districts using a poverty line at the 30th percentile	Percent of districts with same classification	
	Poverty line at 25th percentile	Poverty line at 35th percentile
Poorest group (50 districts)	86%	92%
2nd poorest group (100 districts)	85%	94%
3rd poorest group (100 districts)	80%	95%
Least poor group (293 districts)	95%	99%
Total	90%	97%

Source: Regression analysis of VLSS data at different poverty lines combined with poverty indicators from Agricultural Census.

7. CONCLUSIONS AND DISCUSSION

Poverty maps are useful for geographic targeting of poverty alleviation programs, including government spending on infrastructure, health, education, and nutrition. Disaggregated poverty maps are preferred because the accuracy of geographic targeting has been shown to increase substantially as the size of the geographic unit decreases. This study demonstrates the possibility of combining household survey data and census data to generate highly disaggregated maps of poverty. The method was used to produce a map classifying rural districts in Viet Nam (of which there are over 500), although the same method could be used to classify rural communes (of which there are close to 8,800).

The results also demonstrate that many household characteristics are, individually, fairly weak predictors of poverty. Even indicators such as per capita farm size, per capita food production, and housing material are not very accurate in identifying poor households. Around 60-65 percent of the benefits of a program targeting households by each of these indicators would leak to non-poor households. By contrast, a program distributing benefits randomly would have leakage rates only somewhat higher, 70 percent.

But these household characteristics are more accurate when combined using probit regression analysis. A program targeting households according to a poverty index combining 19 household characteristics and 6 regional dummy variables could reduce leakage and undercoverage rates to 17 percent.

The district-level poverty map suggests that poverty in Vietnam is strongly associated with distance from cities and the coast. It is tempting to suggest that this reflects the economic value of access to markets and that improved transportation infrastructure would reduce poverty in these remote areas. At the same time, it is important to recognize that this spatial pattern may also reflect the influence of other variables that were not included in the analysis. For example, regions that are far from the cities and the coast are more likely to be mountainous and thus less well suited for intensive agriculture.

There are several issues that this paper does not attempt to address, although they are important questions. First, how accurate is this approach of geographic targeting using proxy-indicators and what are the benefits of using it relative to alternative targeting methods? Second, assuming that the poverty map is accurate, how should this information be used to design targeted programs of social spending and infrastructure investment? Third, is the concentration of poverty in the remote areas of the country due to isolation per se, or is it related to other factors correlated with distance such as topography and ethnicity? These questions are left for further research.

REFERENCES

Akerlof, George. 1978. "The economics of 'tagging' as applied to the optimal income tax, welfare programs, and manpower planning." *American Economic Review* 68 (1): 8-19.

Anh, Vu Tuan. 1997. “Criteria for rural poverty monitoring.” *Vietnam’s Socio-Economic Development* 9 (Spring): 3-16.

Baker, Judy and Margaret Grosh. 1994. “Poverty reduction through geographic targeting: How well does it work?" *World Development* 22 (7): 983-995.

Besley, Timothy. 1990. “Means testing versus universal provision in poverty alleviation programmes.” *Economica* 57 (February): 119-129.

Besley, Timothy and Ravi Kanbur. 1993. “The principles of targeting.” In *Including the Poor.* Michael Lipton and Jacque van der Gaag (eds.). Proceedings of a symposium organized by the World Bank and the International Food Policy Research Institute. Regional and Sectoral Studies. The World Bank. Washington, D.C.

Datt, Gaurav and Martin Ravallion. 1993. “Regional disparities, targeting, and poverty in India.” In *Including the Poor.* Michael Lipton and Jacque van der Gaag (eds.). Proceedings of a symposium organized by the World Bank and the International Food Policy Research Institute. Regional and Sectoral Studies. The World Bank. Washington, D.C.

Foster, J., J. Greer, and Eric Thorbecke. 1984. “A class of decomposable poverty measures.” *Econometrica* 52: 761-766.

General Statistical Office. 1995. *Statistical data on basic situation and infrastructure of rural region in Vietnam*. Statistical Publishing House. Hanoi.

Glewwe, Paul. 1990. “Efficient allocation of transfers to the poor: The problem of unobserved household income.” Working Paper No. 70. Living Standards Measurement Study. The World Bank. Washington, D.C.

Grosh, Margaret. 1994. *Administering targeted social programs in Latin America: From Platitudes to Practice*. Regional and Sectoral Studies. The World Bank. Washington, D.C.

Grosh, Margaret and Juan Muñoz. 1996. “A manual for planning and implementing the Living Standards Measurement Study Survey.” Working Paper No. 126. Living Standards Measurement Study. The World Bank. Washington, D.C.

Grosh, Margaret and Judy Baker. 1995. “Proxy means tests for targeting social programs: Simulations and speculation." Working Paper No. 118. Living Standards Measurement Study. The World Bank. Washington, D.C.

Huu, Nguyen Hai. 1997. “Hunger elimination and poverty reduction in the countryside.” *Vietnam’s Socio-Economic Development* 11 (Autumn): 37-46.

Ravallion, Martin. 1993. “Poverty alleviation through regional targeting: A case study for Indonesia.” In *The economics of rural organization*, ed. Karla Hoff, Avery Braverman, and J.E. Stiglitz. Oxford University Press. Oxford.

Ravallion, Martin and Kalvin Chao. 1989. “Targeting policies for poverty alleviation under imperfect information: Algorithms and applications.” *Journal of Policy Modeling* 11 (2): 213-224.

State Planning Committee and General Statistical Office. 1994. *Vietnam Living Standards Survey 1992-93.* Hanoi.

Wodon, Quentin. 1997. “Targeting the poor using ROC curves.” *World Development* 25 (12): 2083-2092.

World Bank. 1995. *Viet Nam: Poverty assessment and strategy*. Washington, D.C.