PART ONE

Application of Tools to Identify the Poor

CHAPTER 1

Predicting Household Poverty Status in Indonesia

Sudarno Sumarto, Daniel Suryadarma, and Asep Suryahadi

Introduction

Indonesia is the fourth most populous country in the world and it has a large poor population. Official poverty estimates indicate that in 2004 the poor numbered about 36 million, or 17 percent of the total population, with about two-thirds of the poor living in rural areas. The most widely used data for measuring poverty is household total consumption expenditure expressed in monetary terms. The use of expenditure data is particularly common in developing countries where expenditure data is less difficult to collect and more accurate than household income data.

Collecting household consumption expenditure data, however, requires plenty of time and effort. Respondents must be willing and patient enough to document their own expenditure over a period of time. For instance, in Indonesia, the recording of food expenditure is done over one week and the enumerators have to ensure that the respondents are correctly noting down their actual expenditure. In addition, some questions on nonfood items require respondents to remember expenditure incurred as far back as one year. In this case, reliability and accuracy of data become an important issue to settle.

Amid such empirical problems, a number of studies in developing countries have been focusing on proxy variables that measure expenditure and poverty. A proxy is calculated using several widely recognized methodologies employing household characteristics data that are auxiliary to poverty and are easier to collect. Examples of proxy variables are asset ownership and education level which can be used to rank households similar to the rank based on per capita consumption expenditure.

One of the more widely cited studies is that of Filmer and Pritchett (1998a), which used long-term household wealth to predict school enrolment in India. The authors employed principal components analysis (PCA) to come up with an asset index for each household. Meanwhile, Ward, Owens, and Kahyrara (2002) and Abeyasekera and Ward (2002) developed proxy predictors of expenditure and income of the poor in Tanzania through the use of the ordinary least squares regression method. A similar study was done by Geda et al. (2001), which uses data from Kenya. Another study is that of Gnawali (2005) that shows the connection between poverty and fertility in Nepal. The Gnawali study employs logistic regression to find out if a household is poor or not by regressing consumption expenditure on some household characteristics. To test the performance of models in predicting welfare, most of these studies compare the rank of households by expenditure with their rank based on the new index developed using PCA.

In most cases, an expenditure variable is used to directly measure poverty, and most studies that employ PCA or the multiple correspondence analysis method to come up with a proxy variable do not exactly aim to estimate expenditure but to capture the multidimensionality of poverty. In a nutshell, this concept argues that poverty does not only involve expenditure or income, but also other dimensions such as health, education, social status, and leisure. Among others, studies that adopt this approach include those of Asselin (2002) and Reyes et al. (2004).

Data and Method

Indonesia's National Socioeconomic Survey (Susenas) data set is used in this study. The Susenas is a nationally representative household survey and has two main components: *core* and *module*. The core component is conducted annually and collects data on household general characteristics and demographic information. The module component contains more detailed characteristics of the households. There are three modules: consumption; health, education, and housing; and social, crime, and tourism. Each module is conducted in turn every year, which means each module is repeated every three years.

Based on a literature study, there are three methods that are commonly used in creating non-income and consumption poverty predictors: (i) by deriving a correlate model of consumption; (ii) by deriving a poverty model with limited dependent variables; and (iii) by calculating a wealth index. In this study, the three methods are explored and compared to get the most appropriate method to determine poverty predictors for Indonesia. Furthermore, since it is widely recognized that conditions in urban and rural areas differ significantly, the best method is implemented separately for urban and rural areas.

Method 1: Consumption Correlate Model

When poverty is defined as a current consumption deficit, a household is categorized as poor if the per capita consumption of its members is lower than a normatively defined poverty line. Therefore, it is logical to search for poverty predictors based on variables that are significantly correlated to per capita household consumption. These variables can be obtained by deriving a correlate model of consumption, where the left-hand side is the per capita consumption while the right-hand side is a set of variables that are thought to be correlated with household consumption. The variables refer to the type of houses and other assets owned by the households, socio-demographic characteristics, and consumption of some specific items. Unlike in the determinant model, in the correlate model the endogeneity of the right-hand side variables is not a concern.[1] (See Appendix 1.1 for the list of the independent variables and their descriptions.)

The dependent variable used is nominal per capita expenditure deflated by implicit deflators for the poverty lines, which vary across provinces to capture the price difference across provinces. Thus, the deflated per capita expenditure is comparable across the country in real terms.

Once the correlates have been determined, the variables are incorporated into the full model and the collinearity of the independent variables to each other is checked. To filter out multicollinearity, a correlation coefficient of each pair of variables is calculated. One of two in a pair of variables is dropped if it is found to be highly correlated and then a regression is run.

Next, a stepwise regression procedure is run to select variables that are appropriate for retention in the model.[2] This procedure facilitates a parsimonious model that has a manageable number of variables but can significantly predict for and explain the variability of household consumption and, hence, poverty status. As this was conducted separately for urban and rural areas, final sets of variables may differ for urban and rural areas.

Finally, in predicting poverty, the performance of the remaining set of variables is tested empirically. For the first step, the variables are used to predict the per capita consumption level of all households in the sample. Second, the predicted per capita consumption is compared with the poverty

1 Take, for example, the car-ownership variable. Generally, one would think that whether a household owns a car or not is determined by, among other factors, its socioeconomic level, and not the other way around. Therefore, car ownership is usually not included in the right-hand side of a consumption determinants model. However, car ownership is a good correlate or predictor of poverty. If a household owns a car, it is most likely that the household is not poor. Hence, this variable should be included in a consumption correlates model.

2 There are three other procedures that can help come up with a parsimonious model, namely, backward, forward, and the all possible regression procedures. The choice is based on the least, but meaningful and practical, number of variables.

line to determine the poverty status of each household. Third, the predicted poverty status is then cross tabulated with the actual poverty status to assess the reliability of the model in predicting poverty. In other words, specificity and sensitivity tests are implemented. A similar test is also conducted to test the reliability of the model in predicting hardcore poverty.[3]

Method 2: Poverty Probability Model

In this model, the dependent variable is a binary variable of household poverty status and the same set (as above) of potential predictor variables is used. The method is known as probit modeling, which is a variant of logit modeling based on different assumptions. Probit may be the more appropriate choice when the categories are assumed to reflect an underlying normal distribution of the dependent variable, even if there are just two categories.[4]

There are two things that need to be reiterated. First, the dependent variable takes the value of *1* when the respondent is poor and *0* when nonpoor. This means that, in interpreting the estimation result, it is important to remember that a positive coefficient means that the variable is correlated positively with the probability of being poor. This is not the case with Method 1, where a positive coefficient means that the variable increases expenditure and hence reduces the chance to be poor. Second, predicted value of the dependent variable is the probability of the observed households being poor. The interpretation of a probit coefficient, say b, is that a one-unit increase in the predictor leads to increasing the probit score by b standard deviations.

Those who prefer to use the first method of using household consumption correlates model to search for poverty predictors argue that a probit model involves unnecessary loss of information in transforming household consumption data into a binary variable. On the other hand, the use of the consumption correlate model to predict poverty also has certain weaknesses. First, estimating a model of consumption correlates does not directly yield a probabilistic statement about household poverty status. Second, the major assumption behind the use of the consumption correlate model is that consumption expenditure is negatively correlated with poverty. Therefore, factors that are found to be positively correlated with consumption are assumed to be automatically negatively correlated with poverty. However, some factors may be positively correlated with consumption but only for

[3] Hardcore poverty is a status of those whose expenditure per capita is below the food poverty line, which means the person cannot satisfy the monthly dietary requirements even when she decides to spend her entire expenditure only on food.

[4] See http://www2.chass.ncsu.edu/garson/pa765/logit.htm for a discussion on this issue.

those who are above the poverty line. However, in general, factors that are positively correlated with welfare are negatively correlated with poverty.

Similarly, a stepwise estimation procedure is also used to produce a manageable number of poverty predictors. As in the first method, specificity and sensitivity tests are also implemented. Total and hardcore poverty are also examined in this method.

Method 3: Wealth Index PCA

One of the indicators of household socioeconomic level is asset ownership. It is relatively easy to collect and can be used to facilitate the wealth ranking of households through the creation of a wealth index. Unfortunately, data on asset ownership is usually in the form of binary variables, indicating only whether a household owns a certain kind of asset or not. Creation of an appropriate wealth index requires data on the quality or price of each asset owned by a household to suitably weigh household assets. Hence, binary data poses a problem in ranking households by their socioeconomic levels.

To deal with this problem, the PCA method is used. In this method, the weight for each asset is determined by the data itself. PCA is a technique for extracting from a large number of variables those few orthogonal linear combinations of the variables that best capture the common information (Filmer and Pritchett 1998b). In effect, it is to reduce the dimensionality (number of variables) of the data set to summarize the most important (i.e., defining), parts while simultaneously filtering out noise. The first principal component is the linear index of variables with the largest amount of information common to all of the variables and each succeeding component accounts for as much of the remaining information as possible. Zeller (2004) stated that the major advantage of PCA is that it does not require a dependent variable (i.e., a household's consumption level or poverty status).

In calculating the PCA index, the method of Filmer and Pritchett (1998b) is adopted:[5]

$$A_j = f_1 \times (a_{j1} - a_1)/(s_1) + \ldots + f_N \times (a_{jN} - a_N)/(s_N) \qquad (1)$$

or simply

$$A_j = \sum_{i=1}^{N} \frac{f_i (a_{ji} - a_i)}{s_i}$$

5 They refer to it as Economic Status Index. Although Filmer and Pritchett (1998a, 1998b) cautioned that they are not proposing the wealth index be used as a proxy for current living standards or poverty analysis, they tested the index's robustness using current consumption expenditures and poverty rates data. Thus, if the index is as robust as they claimed, then it would not be a problem to use it as a proxy for current living standards.

where

f_i is the 'scoring factor' for the i^{th} asset determined by the method

a_{ji} is the jth household's value for the i^{th} asset and

a_{ji} and s_i are the mean and standard deviation respectively of the i^{th} asset variable over all households

A_j = Asset index of the jth household.

Note that the mean value of the index is zero by construction since it is a weighted sum of the mean deviations. Based on the results of this analysis, households can be ranked from the lowest to the highest socioeconomic level. Testing the reliability of this wealth ranking on predicting poverty requires a cutoff point to separate the predicted poor from the nonpoor. Since there is no a priori poverty line that can be determined objectively in the PCA method, the cutoff point used is determined such that the poverty ratio predicted by the PCA method is the same as that derived from the actual consumption expenditure distribution. The additional value added from the PCA method lies in easy identification of the poor households through an asset index even when the overall percentage of poor might be the same as when PCA and consumption expenditure methods are used.

As in the first two methods, a cross tabulation is performed between the results of this approach and the poverty status based on the actual consumption expenditure.

The Poverty Line

The poverty line and food poverty line of Indonesia used in this study are the ones calculated by Pradhan et al. (2001). The food poverty line is based on a single national bundle of food producing 2,100 calories per person a day priced by nominal regional prices. This means that the differences in the value of this food poverty line across regions arise solely from price differences across regions. The nonfood poverty line component is estimated using the Engel law method. The total and food poverty lines used in this study are shown in Appendix 1.2.

Results

Correlate Model Method

When checking for the presence of multicollinearity, correlation coefficients of the final set of variables generated are found to be not higher than 0.7–implying the multicollinearity issue has been minimized. After running the stepwise procedure, the retained variables in the model (Table 1.1), provide R-squared equal to 44 percent. This result means that these variables can explain 44 percent variability in per capita consumption of urban households and 36 percent variability of rural households. The result is close to that in Ward, Owens, and Kahyrara (2002) where around 40 percent of variation is explained. Furthermore, most of the coefficients have signs as expected. However, the set of significant variables in urban areas is not the same as that in rural areas. In addition, as discussed below, the coefficients of some variables have opposite signs in urban and rural areas (See Appendix 1.3 for details).

Table 1.1 **Summary Results of Ordinary Least Squares Regression of the Consumption Correlates Model**

Item	Urban	Rural
Number of observations	23,847	34,649
Adjusted R-squared	0.44	0.36

Source: Authors' calculation based on 2004 SUSENAS.

Coefficients of the asset-ownership group of variables for urban areas are all positive, indicating that ownership of these various assets is correlated with a higher level of household welfare. In both urban and rural areas, the ownership of a car, refrigerator, motorcycle, and satellite dish are the variables with the highest correlations with consumption. Interestingly, households which raise chickens in rural areas have higher per capita consumption than those that do not, but raising chickens in urban areas is negatively correlated with per capita consumption.

Like asset ownership, the coefficients for household characteristics variables indicate that better housing materials are correlated with higher per capita consumption. In urban areas, a tile roof and a concrete wall are the two household characteristics that have the highest correlation coefficients with consumption, while the highest coefficients in rural areas are observed for households with an electrical connection to the house and flush toilets.

The correlation coefficients of variable age with consumption also differ in urban and rural areas. In rural areas, the age of the household head has a significant positive relationship. On the other hand, in urban areas, it is the age of the household spouse that has a significant, but negative, relationship.

The education level of the household head is a strong predictor of per capita consumption in both urban and rural areas. The higher the education level of the household head, the higher the per capita consumption. However, the marginal impact of each education level on consumption is much higher in urban areas than in rural areas.

In addition, the education level of a spouse is negatively correlated with consumption. This is an unexpected and puzzling result in both urban and rural areas. The marginal impact of each education level on consumption is also much higher in urban areas than in rural areas. In interpreting this negative correlation, it has to be remembered that the correlations are controlled by holding other variables constant. One possibility is that these negative coefficients may indicate that, all other things being equal, households with spouses that have higher education levels save more, hence they consume less.

In rural areas, the enrollment status of school-age children is also significantly related with consumption. In these areas, households which have at least one child aged 6–15 years who has dropped out of school have significantly lower per capita consumption.

In both urban and rural areas, larger household size is correlated with lower per capita consumption. The coefficients of the squared household-size variable indicate that the reduction in per capita consumption as household size gets larger occurs at a decreasing rate. Furthermore, higher dependency ratio–defined as the proportion of household members aged less than 15 years–of a household is also correlated with lower per capita consumption.

The working status of a spouse is positively correlated with per capita consumption. However, this correlation is only statistically significant for urban areas. Likewise, households which have children aged 6–15 years who are working also have higher per capita consumption and this is true in both urban and rural areas. In rural areas, having a household head working in the formal sector is also positively correlated with per capita consumption.

In both urban and rural areas, clothing turns out to have a strong correlation with consumption. Households in which each member has different clothing for different activities have higher per capita consumption. In rural areas, the use of modern medicine for curing sickness is also positively associated with per capita consumption.

Finally, the pattern of consumption itself is a strong predictor of the level of consumption. In urban areas, households in which each member eats at least twice a day have higher per capita consumption. Moreover, in both urban and rural areas, households that consume beef, eggs, milk, biscuits, bread, and bananas at least once in a week have higher per capita consumption. On the other hand, households in rural areas which consume *tiwul* (cassava flour), an inferior good, at least once a week have lower per capita consumption.

These estimation results are then used to predict per capita consumption of households given their characteristics. The accuracy of this predicted consumption is examined by cross tabulating it with actual consumption, where both the predicted and actual consumption are ranked and divided into three groups: bottom 30 percent, middle 40 percent, and top 30 percent. Table 1.2 shows the results of the cross tabulation for both urban and rural areas. If the household grouping based on predicted consumption perfectly matches the grouping by actual consumption, then all the diagonal cells will be 100 percent and off-diagonal cells will be 0.

Table 1.2 **Accuracy of Predicting Expenditure Using the Consumption Correlates Model**

Percentage (%) of Urban Consumption Expenditure

		Predicted		
		Bottom 30%	*Middle 40%*	*Top 30%*
Actual	*Bottom 30%*	67.33	30.22	2.45
	Middle 40%	22.44	56.57	20.99
	Top 30%	2.75	27.67	69.57

Percentage (%) of Rural Consumption Expenditure

		Predicted		
		Bottom 30%	*Middle 40%*	*Top 30%*
Actual	*Bottom 30%*	63.40	32.18	4.42
	Middle 40%	24.14	53.42	22.44
	Top 30%	4.41	29.93	65.67

Source: Authors' calculation.

In urban areas, 67.3 percent of households are correctly predicted to be in the bottom 30 percent, while only 2.5 percent of those households are wrongly predicted to be in the top 30 percent. Meanwhile, for those who are actually in the top 30 percent, 69.6 percent are predicted correctly, while about 2.7 percent are wrongly predicted to be in the bottom 30 percent. For the 40 percent in the middle, 56.6 percent are accurately predicted, while the remaining 43.0 percent are predicted almost equally split to be in the top or bottom 30 percent.

In rural areas, about 63.4 percent of people in the bottom 30 percent are predicted correctly, while 4.4 percent are wrongly predicted to be in the top 30 percent. On the other hand, 65.7 percent of those in the top 30 percent are accurately predicted and also 4.4 percent are wrongly predicted to be in the top 30 percent. Meanwhile, 53.4 percent of the middle group households are predicted to be where they are.

On an average, 64.5 percent of households' position in the per capita consumption groups is predicted correctly in urban areas and 60.8 percent in rural areas. As expected, prediction in urban areas is more accurate because of the higher coefficient of determination in the regression results.

Next, the accuracy of the model in predicting poverty is examined. Since poverty lines have been previously defined, the households with predicted expenditure below the poverty line are considered poor. Table 1.3 shows the result for poverty and Table 1.4 for hardcore poverty. Since the interest is in predicting poverty, the accuracy of predicting the nonpoor is less relevant. As shown in Table 1.3, in urban areas, around 49.6 percent of the poor are correctly predicted as poor; the result is slightly lower in rural areas, where 45.7 percent are correctly predicted. This indicates that predicted expenditure tends to underestimate poverty. Therefore, if predicted expenditure is used as a targeting tool for the poor in urban areas, there will be under-coverage of 50.4 percent for the share of poor who are wrongly predicted to be nonpoor, and about 7.3 percent of the nonpoor will benefit from the program.

Table 1.3 **Accuracy of Predicting Poverty Using the Consumption Correlates Model**

Percentage of Urban Poverty			
		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	92.73	7.27
	Poor	50.43	49.57

Percentage of Rural Poverty			
		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	92.12	7.88
	Poor	54.32	45.68

Source: Authors' calculation.

Meanwhile, Table 1.4 shows that the prediction results are even lower for hardcore poverty. Around 48.4 percent of the hardcore poor in urban areas and 33.5 percent of the hardcore poor in rural areas are correctly classified.

In conclusion, Method 1 produces quite robust results and is relatively accurate when used to predict consumption expenditure. However, the method performs less well when used to predict poverty as only around one half of the poor are predicted correctly.

Table 1.4 **Accuracy of Predicting Hardcore Poverty Using the Consumption Correlates Model**

Percentage of Urban Poverty			
		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	94.62	5.38
	Poor	51.55	48.45

Percentage of Rural Poverty			
		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	95.60	4.40
	Poor	66.52	33.48

Source: Authors' calculation.

Poverty Probability Method

The poverty probability method predicts poverty directly because of the nature of the dependent variable. The result of the poverty estimation for Indonesia is in Table 1.5, while the result of hardcore poverty estimation is in Table 1.6.

For the poverty estimation, the pseudo R-squared is 0.36 for urban areas and 0.29 for rural areas. For hardcore poverty estimation, the pseudo R-squared is 0.35 for urban and 0.28 for rural areas. In general, the coefficients in the results of the poverty probability model (Table 1.5) are consistent with those in the ordinary least squares regression results of the consumption correlates model (Table 1.4). For example, the asset ownership variables have positive coefficients in Table 1.4 which means that households that own various assets are more likely to have higher consumption expenditures. Meanwhile, in the results of the poverty probability model (Table 1.5), the coefficients of these asset ownership variables are negative, which means that households that own various assets are less likely to be poor. These results are hence consistent with each other.

There are, however, some exceptions. For example, in Table 1.4 the variable of owning a sewing machine is dropped as a result of stepwise regression in both urban and rural areas, implying that owning a sewing machine is not correlated significantly with the level of household per capita consumption. However, in Table 1.5 the coefficient of this variable is negative and significant for rural areas, which means that rural households that own sewing machines have a lower probability of being poor.

Furthermore, it is interesting to see the difference between poverty predictors and hardcore poverty predictors. Table 1.6 reveals that after implementing a stepwise procedure, fewer significant predictors for the hardcore poor are retained compared with those for the poor. For instance, the results indicate that relative to households with heads having education less than primary level, the higher the education level of the household head, the lower the probability of that the household is poor. For the hardcore poor, results indicate that only households whose heads are at least graduates from senior high school have significant lower probability of being hardcore poor.

The accuracy of predicting actual poverty using Method 2 can also be observed. The predicted value of the dependent variable is the probability of households to be poor given their characteristics. To classify households into predicted poor and predicted nonpoor, we need a threshold to separate these two groups of households. Following Pritchett, Suryahadi, and Sumarto

Table 1.5 Results of the Poverty Probability Model
(Dependent Variable: 1 = Poor, 0 = Otherwise)

Predictors	Urban Areas	Rural Areas
Asset Ownership		
this household owns a sewing machine		-0.118**
		[0.033]
this household owns a radio	-0.110**	-0.130**
	[0.030]	[0.018]
this household owns a television	-0.243**	-0.171**
	[0.032]	[0.022]
this household owns a refrigerator	-0.408**	-0.319**
	[0.051]	[0.063]
this household owns jewelry	-0.225**	-0.223**
	[0.028]	[0.019]
this household owns a satellite dish		-0.291**
		[0.071]
this household owns a bicycle or a boat		-0.159**
		[0.019]
this household owns a motorcycle	-0.544**	-0.471**
	[0.041]	[0.030]
this household owns a car	-0.488**	-0.380**
	[0.104]	[0.083]
Animal Ownership		
this household owns a cow		0.065**
		[0.022]
this household owns a chicken		-0.106**
		[0.017]
this household owns other animal	0.403**	
	[0.141]	
House Characteristics		
wall of the house is made from concrete	-0.206**	-0.137**
	[0.032]	[0.021]
floor of the house is dirt floor	0.214**	0.144**
	[0.049]	[0.023]
toilet type of the house is flush	-0.220**	-0.133**
	[0.031]	[0.023]
this household uses its own toilet	-0.105**	
	[0.032]	
this household has electricity	-0.232**	-0.194**
	[0.060]	[0.022]
this household's source of water is from protected well or water pump	-0.231**	-0.150**
	[0.036]	[0.019]
Household Characteristics		
household head age	-0.035**	-0.033**
	[0.006]	[0.004]
household head age squared	0.000**	0.000**
	[0.000]	[0.000]
spouse age		-0.002**
		[0.001]
household head finishes primary education	-0.111**	-0.082**
	[0.034]	[0.021]
household head finishes junior secondary education	-0.210**	-0.134**
	[0.043]	[0.034]
household head finishes senior secondary education	-0.271**	-0.245**
	[0.044]	[0.041]
household head finishes tertiary education	-0.640**	-0.517**
	[0.104]	[0.126]
spouse finishes primary education		0.087**
		[0.021]
household size	0.627**	0.649**
	[0.028]	[0.021]

(continued on next page)

Table 1.5 continued

Predictors	Urban Areas	Rural Areas
household size squared	-0.030**	-0.032**
	[0.002]	[0.002]
dependency ratio of this household is more than 0.5	0.284**	0.200**
	[0.041]	[0.027]
household head is working		-0.119**
		[0.036]
spouse is working	-0.110**	
	[0.028]	
household head is working in the formal sector		-0.099**
		[0.026]
at least one school-age child (6–15 years old) in this household has dropped out of school	0.172**	0.122**
	[0.042]	[0.025]
at least one school-age child (6–15 years old) in this household is working		-0.098**
		[0.033]
main source of income for this household is from agricultural sector	0.143**	0.094**
	[0.037]	[0.022]
every household member has different clothing for different activities	-0.295**	-0.389**
	[0.065]	[0.040]
when a member in this household is sick, s/he is treated with modern medicine		-0.113**
		[0.027]
Consumption Pattern		
this household consumed beef in the past week	-0.346**	-0.405**
	[0.056]	[0.053]
this household consumed egg in the past week	-0.328**	-0.325**
	[0.027]	[0.019]
this household consumed milk in the past week	-0.573**	-0.644**
	[0.047]	[0.045]
this household consumed biscuit in the past week	-0.207**	-0.205**
	[0.045]	[0.031]
consumed bread in the past week	-0.209**	-0.221**
	[0.032]	[0.022]
this household consumed banana in the past week	-0.139**	-0.291**
	[0.040]	[0.026]
this household consumed *tiwul* in the past week		0.162**
		[0.055]
Constant	-1.432**	0.172
	[0.174]	[0.107]
Province dummy variables included	Yes	Yes
Number of observations	23,847	34,649
Pseudo R-squared	0.362	0.288

** Significant at 1%; * Significant at 5%
[ ] Robust standard errors in bracket
Source: Authors' calculation based on 2002 SUSENAS.

(2000) and Suryahadi and Sumarto (2003a and 2003b), we use a 50 percent probability of being poor as the threshold. Hence, households which have 50 percent or higher probability to be poor are classified as predicted poor, while households which have less than fair probability to be poor are classified as predicted nonpoor. Using this 50 percent probability threshold, Tables 1.7 and 1.8 show, respectively, the cross tabulations between the actual and predicted poverty conditions.

Table 1.6 **Results of the Poverty Probability Model**
(Dependent Variable: 1= Hardcore Poor, 0 = Otherwise)

Predictors	Urban Areas	Rural Areas
Asset Ownership		
this household owns a sewing machine		-0.135**
		[0.044]
this household owns a radio	-0.124**	-0.152**
	[0.042]	[0.022]
this household owns a television	-0.322**	-0.159**
	[0.044]	[0.027]
this household owns a refrigerator	-0.332**	-0.305**
	[0.088]	[0.092]
this household owns jewelry	-0.213**	-0.248**
	[0.040]	[0.023]
this household owns a satellite dish		-0.448**
		[0.111]
this household owns a bicycle or a boat		-0.175**
		[0.023]
this household owns a motorcycle	-0.315**	-0.413**
	[0.064]	[0.042]
this household owns a car	-0.682**	
	[0.236]	
Animal Ownership		
this household owns a chicken		-0.101**
		[0.021]
House Characteristics		
wall of the house is made from concrete	-0.286**	-0.166**
	[0.043]	[0.026]
floor of the house is dirt floor		0.135**
		[0.026]
toilet type of the house is flush	-0.189**	
	[0.045]	
this household uses its own toilet	-0.148**	
	[0.045]	
this household has electricity		-0.237**
		[0.025]
this household's source of water is from protected well or water pump	-0.168**	-0.149**
	[0.047]	[0.022]
Household Characteristics		
household head age	-0.028**	-0.032**
	[0.008]	[0.005]
household head age squared	0.000**	0.000**
	[0.000]	[0.000]
spouse age		-0.002**
		[0.001]
household head finishes senior secondary education	-0.283**	-0.165**
	[0.066]	[0.052]
household head finishes tertiary education	-0.960**	
	[0.287]	
spouse finishes primary education		0.066**
		[0.023]
household size	0.509**	0.590**
	[0.039]	[0.023]
household size squared	-0.022**	-0.028**
	[0.003]	[0.002]
dependency ratio of this household is more than 0.5	0.325**	0.165**
	[0.053]	[0.030]
household head is working		-0.180**
		[0.042]
household head is working in the formal sector		-0.180**
		[0.033]

(continued on next page)

Table 1.6 continued

Predictors	**Urban Areas**	**Rural Areas**
at least one school-age child (6–15 years old) in this household has dropped out of school	0.141**	0.116**
	[0.052]	[0.026]
main source of income for this household is from agricultural sector	0.138**	0.101**
	[0.048]	[0.027]
every household member has different clothing for different activities	-0.382**	-0.366**
	[0.081]	[0.042]
when a member in this household is sick, s/he is treated with modern medicine		-0.152**
		[0.032]
Consumption Pattern		
every household member eats at least twice a day	-0.452**	-0.276**
	[0.118]	[0.073]
this household consumed beef in the past week	-0.455**	-0.494**
	[0.094]	[0.070]
this household consumed egg in the past week	-0.414**	-0.416**
	[0.040]	[0.025]
this household consumed milk in the past week	-0.627**	-0.689**
	[0.085]	[0.067]
this household consumed biscuit in the past week		-0.210**
		[0.040]
this household consumed bread in the past week	-0.249**	-0.195**
	[0.048]	[0.028]
this household consumed banana in the past week		-0.301**
		[0.034]
this household consumed *tiwul* in the past week		0.185**
		[0.057]
Constant	-1.506**	-0.081
	[0.231]	[0.140]
Province dummy variables included	Yes	Yes
Observations	23759	34649
Pseudo R-squared	0.352	0.28

** Significant at 1%; * Significant at 5%
[ ] Robust standard errors in bracket
Source: Authors' calculation based on 2002 SUSENAS.

Table 1.7 shows that 35.6 percent of the poor are predicted correctly in urban areas and less than 3.0 percent of the nonpoor are predicted to be poor. Meanwhile, in rural areas about 52.7 percent of the poor are predicted correctly, even though the percentage of the nonpoor predicted to be poor is also higher, 9.5 percent.[6] Prediction for urban areas is much less accurate than using Method 1, where almost 50 percent of the poor are correctly predicted. However, the prediction in rural areas is better than when using Method 1.

Table 1.8 shows that predicted hardcore poverty is even less accurate than predicted poverty. Comparing Table 1.8 with Table 1.4, Method 2 makes worse predictions than Method 1. Thus, the only instance where prediction

6 The authors readily admit that changing the 50 percent threshold of poverty probability will also change the accuracy. For example, by using 30 percent as the threshold, we get higher accuracy. However, using less than 50 percent as a threshold is hard to justify, thus, the authors opt to use the 50 percent threshold, which implies even chances for poor and nonpoor.

is better when using Method 2 than Method 1 is for predictions of poverty in rural areas.

Wealth Index PCA Method

Table 1.9 provides the scoring factor, mean, and standard deviation of each variable for urban areas, while Table 1.10 provides those for rural areas. The mean of the indexes in both areas are zero by construction.

The fifth column, scoring factor/ standard deviation, is the increase in the wealth index if the household moves from *0* to *1* on a dummy variable. For example, a household in urban areas will increase its wealth index by 0.71 if it owns a car. Car ownership has the highest score, while living in a dirt-floor residence has the most negative score. For rural areas, the highest score is obtained with a spouse having a tertiary education, which increases the index by 1.1, and the lowest score is if the household is in the agricultural sector, which dropped the index to -0.47.

Table 1.7 **Accuracy of Predicting Poverty Using the Poverty Probability Model**

Percentage of Urban Poverty

		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	97.07	2.93
	Poor	64.44	35.56

Percentage of Rural Poverty

		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	90.49	9.51
	Poor	47.33	52.67

Source: Authors' calculation.

Table 1.8 **Accuracy of Predicting Hardcore Poverty Using the Poverty Probability Model**

Percentage of Urban Poverty

		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	99.66	0.34
	Poor	87.89	12.11

Percentage of Rural Poverty

		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	97.62	2.38
	Poor	73.67	26.33

Source: Authors' calculation.

Table 1.11 shows a cross tabulation between terciles of households based on the wealth index as a measure of predicted consumption expenditure and terciles of households based on actual per capita consumption expenditure for urban and rural areas. In urban areas, 51.1 percent of those in the bottom 30 percent and 54.6 percent of those in the top 30 percent are predicted correctly using Method 3. On the other hand, in rural areas 47.4 percent of those in the bottom 30 percent and 50.3 percent of those in the top 30 percent are accurately predicted. The accuracy of this approach is much lower than that achieved by Method 1, where more than 60 percent of each tercile is predicted correctly.

To measure the performance of this approach in predicting poverty, a threshold is needed to divide households into those that are predicted as poor and those predicted as nonpoor. Since there is no such threshold in the wealth index that can be calculated objectively, it is assumed that the

Table 1.9 **Summary Statistics and Eigen-value**
(First Principal Component), Urban Area

Predictors	**Scoring Factor**	**Mean**	**Standard Deviation**	**Scoring Factor / Std Dev**
this household owns a sewing machine	0.175	0.253	0.435	0.40
this household owns a radio	0.208	0.781	0.413	0.50
this household owns a television	0.286	0.729	0.445	0.64
this household owns a refrigerator	0.305	0.303	0.460	0.66
this household owns jewelry	0.226	0.604	0.489	0.46
this household owns a satellite dish	0.178	0.111	0.314	0.57
this household owns a bicycle or a boat	0.083	0.401	0.490	0.17
this household owns a motorcycle	0.233	0.294	0.456	0.51
this household owns a car	0.200	0.086	0.280	0.71
this household owns land	0.015	0.264	0.441	0.03
this household owns the house they're living in	0.038	0.871	0.335	0.11
roof of the house is made from tile	0.034	0.618	0.486	0.07
wall of the house is made from concrete	0.173	0.701	0.458	0.38
floor of the house is dirt floor	-0.149	0.046	0.210	-0.71
toilet type of the house is flush	0.235	0.702	0.457	0.51
this household uses its own toilet	0.251	0.697	0.460	0.55
this household has electricity	0.139	0.968	0.176	0.79
this household's source of water is from protected well or water pump	0.115	0.867	0.340	0.34
this household owns a cow	-0.055	0.019	0.137	-0.40
this household owns a goat	-0.048	0.019	0.135	-0.35
this household owns chicken	-0.053	0.152	0.359	-0.15
this household owns other animal	-0.009	0.005	0.074	-0.12
household head age	-0.001	44.740	13.639	0.00
spouse age	0.138	31.580	18.389	0.01
household head finishes primary education	-0.105	0.247	0.431	-0.24
household head finishes junior secondary education	-0.005	0.165	0.371	-0.01
household head finishes senior secondary education	0.138	0.290	0.454	0.30
household head finishes tertiary education	0.180	0.097	0.297	0.61
spouse finishes primary education	-0.050	0.240	0.427	-0.12
spouse finishes junior secondary education	0.055	0.144	0.351	0.16
spouse finishes senior secondary education	0.184	0.194	0.395	0.47
spouse finishes tertiary education	0.139	0.048	0.214	0.65
household size	0.128	4.335	1.870	0.07
dependency ratio of this household is more than 0.5	0.001	0.092	0.289	0.00
household head is working	0.056	0.846	0.361	0.15
spouse is working	0.073	0.352	0.478	0.15
household head is married	0.144	0.829	0.376	0.38
household head is working in formal sector	0.176	0.535	0.499	0.35
at least one school-age child (6–15 years old) in this household has dropped out of school	-0.054	0.077	0.266	-0.20
at least one school-age child (6–15 years old) in this household is working	-0.022	0.025	0.156	-0.14
main source of income for this household is from agricultural sector	-0.136	0.093	0.290	-0.47
every household member eats at least twice a day	0.024	0.987	0.113	0.21
every household member has different clothing for different activities	0.083	0.974	0.161	0.52
when a member in this household is sick, s/he is treated with modern medicine	0.091	0.926	0.262	0.35
this household consumed *gaplek* in the past week	-0.003	0.004	0.061	-0.05
this household consumed *tiwul* in the past week	-0.007	0.001	0.033	-0.21
this household consumed beef in the past week	0.159	0.147	0.354	0.45
this household consumed egg in the past week	0.143	0.634	0.482	0.30
this household consumed milk in the past week	0.188	0.247	0.431	0.44
this household consumed biscuit in the past week	0.072	0.130	0.336	0.21
this household consumed bread in the past week	0.075	0.280	0.449	0.17
this household consumed banana in the past week	0.089	0.180	0.384	0.23
PCA Index		0.000	2.207	

Std dev = standard deviation
Source: Authors' calculation.

Table 1.10 **Summary Statistics and Eigen-value**
(First Principal Component), Rural Area

Predictors	Scoring Factor	Mean	Standard Deviation	Scoring Factor / Std Dev
this household owns a sewing machine	0.174	0.123	0.329	0.53
this household owns a radio	0.202	0.603	0.489	0.41
this household owns a television	0.301	0.377	0.485	0.62
this household owns a refrigerator	0.214	0.050	0.218	0.98
this household owns jewelry	0.202	0.463	0.499	0.41
this household owns a satellite dish	0.183	0.046	0.209	0.88
this household owns a bicycle or a boat	0.118	0.426	0.494	0.24
this household owns a motorcycle	0.240	0.163	0.369	0.65
this household owns a car	0.131	0.025	0.156	0.84
this household owns land	-0.062	0.722	0.448	-0.14
this household owns the house they're living in	-0.004	0.945	0.228	-0.02
roof of the house is made from tile	0.060	0.591	0.492	0.12
wall of the house is made from concrete	0.213	0.419	0.493	0.43
floor of the house is dirt floor	-0.164	0.217	0.412	-0.40
toilet type of the house is flush	0.269	0.264	0.441	0.61
this household uses its own toilet	0.1914	0.447	0.497	0.38
this household has electricity	0.216	0.736	0.441	0.49
this household's source of water is from protected well or water pump	0.168	0.504	0.500	0.34
this household owns a cow	-0.066	0.179	0.384	-0.17
this household owns a goat	-0.049	0.114	0.318	-0.16
this household owns a chicken	-0.035	0.465	0.499	-0.07
this household owns other animal	-0.013	0.014	0.117	-0.11
household head age	-0.072	45.905	14.043	-0.01
spouse age	0.069	32.770	18.249	0.00
household head finishes primary education	-0.003	0.339	0.474	-0.01
household head finishes junior secondary education	0.073	0.094	0.292	0.25
household head finishes senior secondary education	0.185	0.095	0.293	0.63
household head finishes tertiary education	0.140	0.019	0.136	1.03
spouse finishes primary education	0.039	0.300	0.458	0.09
spouse finishes junior secondary education	0.099	0.072	0.258	0.38
spouse finishes senior secondary education	0.170	0.055	0.228	0.75
spouse finishes tertiary education	0.108	0.010	0.098	1.10
household size	0.073	4.129	1.759	0.04
dependency ratio of this household is more than 0.5	-0.014	0.113	0.317	-0.05
household head is working	0.040	0.923	0.267	0.15
spouse is working	0.028	0.501	0.500	0.06
household head is married	0.115	0.855	0.352	0.33
household head is working in the formal sector	0.232	0.239	0.426	0.54
at least one school-age child (6–15 years old) in this household has dropped out of school	-0.072	0.148	0.355	-0.20
at least one school-age child (6–15 years old) in this household is working	-0.053	0.068	0.251	-0.21
main source of income for this household is from agricultural sector	-0.222	0.596	0.491	-0.45
every household member eats at least twice a day	0.029	0.986	0.116	0.25
every household member has different clothing for different activities	0.084	0.962	0.192	0.44
when a member in this household is sick, s/he is treated with modern medicine	0.108	0.892	0.311	0.35
this household consumed *gaplek* in the past week	-0.030	0.012	0.107	-0.28
this household consumed *tiwul* in the past week	-0.038	0.021	0.144	-0.26
this household consumed beef in the past week	0.118	0.048	0.215	0.55
this household consumed egg in the past week	0.163	0.368	0.482	0.34
this household consumed milk in the past week	0.169	0.088	0.283	0.60
this household consumed biscuit in the past week	0.072	0.103	0.303	0.24
this household consumed bread in the past week	0.077	0.208	0.406	0.19
this household consumed banana in the past week	0.054	0.144	0.351	0.15
PCA Index		0.000	2.180	

Std dev = standard deviation
Source: Authors' calculation.

Table 1.11 **Accuracy of Predicting Per Capita Consumption Expenditure Using the Wealth Index Principal Component Analysis**

Percentage of Urban Consumption Expenditure

		Predicted based on wealth index		
		Bottom 30%	*Middle 40%*	*Top 30%*
Actual	*Bottom 30%*	51.10	41.52	7.38
	Middle 40%	25.79	45.69	28.52
	Top 30%	14.51	30.89	54.61

Percentage of Rural Consumption Expenditure

		Predicted based on wealth index		
		Bottom 30%	*Middle 40%*	*Top 30%*
Actual	*Bottom 30%*	47.35	40.73	11.92
	Middle 40%	26.84	44.78	28.38
	Top 30%	16.85	32.90	50.25

Source: Authors' calculation.

threshold is the value of the wealth index at the percentile of the actual poverty rate. For example, if the poverty rate is X percent, then the threshold is the value of the wealth index at the X^{th} percentile. In other words, this is the threshold which will result in X percent predicted poverty rate, which is the same as the actual poverty rate. Using this threshold, Tables 1.12 and 1.13 show the cross tabulation between actual and predicted rates for poverty and hardcore poverty, respectively.

Table 1.12 **Accuracy of Predicting Poverty Using the Wealth Index Principal Component Analysis**

Percentage of Urban Poverty

		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	90.14	9.86
	Poor	64.72	35.28

Percentage of Rural Poverty

		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	78.12	21.88
	Poor	53.68	46.32

Source: Authors' calculation.

Table 1.12 reveals that only 35.3 percent of the poor in urban areas are predicted correctly, making the wealth index PCA the least accurate of the three approaches for predicting poverty. However, 46.3 percent of poor people in rural areas are predicted correctly, which is a higher rate than when Method 1 is used (45.7 percent) but lower when Method 2 is used (52.7 percent).

Table 1.13 **Accuracy of Predicting Hardcore Poverty Using the Wealth Index Principal Component Analysis**

Percentage of Urban Poverty

		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	96.43	3.57
	Poor	81.68	18.32

Percentage of Rural Poverty

		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	89.20	10.80
	Poor	68.14	31.86

Source: Authors' calculation.

Meanwhile, in predicting hardcore poverty, 31.9 percent of the hardcore poor in rural areas and 18.3 percent in urban areas are predicted correctly when the wealth index PCA is used (Table 1.13). Compared with the performance of the other approaches in predicting hardcore poverty, the accuracy of this approach is higher than Method 2 but lower than Method 1.

Conclusion

In the face of the difficulties in acquiring household expenditure and income data, three methods for predicting poverty were explored in this study. These three approaches were the consumption correlates model, poverty probability model, and wealth index PCA. In terms of predicting expenditure, the consumption correlates model is the best approach as it is able to predict correctly the poverty status of more than 60 percent of the respondents in both urban and rural areas.

In terms of predicting poverty and hardcore poverty, the results were mixed. In hardcore poverty prediction, the best approach was by far the consumption correlates model. In predicting poverty, the poverty probability model was the best predictor for rural areas (52.7 percent accurate), while for urban areas the consumption correlates model provided the best result (49.6 percent accurate). In conclusion, the consumption model is, all things being equal, the best approach to be used to find expenditure and poverty predictors.

A common thread in the predictions is that the better poverty prediction is, the more nonpoor are predicted to be poor. Thus, the method that makes the most accurate prediction, also predicts the most nonpoor to be poor.

Furthermore, empirical results show that variables with the strongest correlates, negative or positive, are car and refrigerator ownership, education level, household size, and consumption of milk and beef. In addition, playing relatively small but significant roles are house characteristics, access to facilities, and employment status of household members. Thus, for a rough assessment on whether a household is more likely to be poor or not in Indonesia, it would be best to gather information on asset ownership, education level, and consumption patterns.

Further avenues of research on this subject include finding methods to take into account the quality or prices of assets owned or food consumed, since quality can also distinguish nonnegligibly between poor and nonpoor households.

Appendix

Appendix 1.1 List of Variables Used to Estimate Expenditure and Poverty Predictors

Group	Variable	Description
Asset	own_sewing machine	this household owns a sewing machine
	own_radio	this household owns a radio
	own_tv	this household owns a television
	own_fridge	this household owns a refrigerator
	own_jewelry	this household owns jewelry
	own_satdish	this household owns a satellite dish
	own_bikeboat	this household owns a bicycle or a boat
	own_motorcycle	this household owns a motorcycle
	own_car	this household owns a car
	own_land	this household owns land
	own_house	this household owns the house they are living in
House	tile roof	roof of the house is made from tile
	concrete wall	wall of the house is made from concrete
	dirtfloor	floor of the house is made from dirt
	flushtoilet	toilet type of the house is flush
	own_toilet	this household uses its own toilet
	electric_light	this household has electricity
	protectedwatersrc	this household's source of water is from protected well or water pump
Farm	own_cow	this household owns a cow
	own_goat	this household owns a goat
	own_chicks	this household owns a chicken
	own_othanim	this household owns other animal
Household	age	household head age
	spage	spouse age
	elm	household head finishes primary education
	lsec	household head finishes junior secondary education
	usec	household head finishes senior secondary education
	ter	household head finishes tertiary education
	spelm	spouse finishes primary education
	splsec	spouse finishes junior secondary education
	spusec	spouse finishes senior secondary education
	spter	spouse finishes tertiary education
	fsize	household size
	deprhigh	dependency ratio of this household is more than 0.5
	headwork	household head is working
	spwork	spouse is working
	marr	household head is married
	formal	household head is working in the formal sector
	child_dropout	at least one school-age child (6–15 years old) in this household has dropped out of school
	child_work	at least one school-age child (6–15 years old) in this household is working
	in_agric	main source of income for this household is from agricultural sector
	eattwice	every household member eats at least twice a day
	clothes	every household member has different clothing for different activities
	usemodernmed	when a member in this household is sick, s/he is treated with modern medicine
Consumption	cgaplek	this household consumed *gaplek* (dried cassava) in the past week
	ctiwul	this household consumed *tiwul* (cassava flour) in the past week
	cbeef	this household consumed beef in the past week
	cegg	this household consumed *egg* in the past week
	cmilk	this household consumed milk in the past week
	cbiscuit	this household consumed biscuit in the past week
	cbread	this household consumed bread in the past week
	cbanana	this household consumed banana in the past week

Note: Variables are binary (0/1) variables, except age, spage, fsize.
Source: Authors' calculation based on 2002 SUSENAS.

Appendix 1.2 **Poverty Lines in February 1999**
(Rp per capita per month)

Province	Poverty Line		Food Poverty Line	
	Urban	Rural	Urban	Rural
Nanggroe Aceh Darussalam	74,064	70,280	60,733	60,003
North Sumatera	83,745	74,712	66,803	63,753
West Sumatera	85,409	78,762	69,668	66,416
Riau	92,970	82,420	73,812	70,654
Jambi	85,874	77,104	68,078	65,841
South Sumatera	86,154	80,033	68,830	67,585
Bengkulu	86,714	77,750	67,958	64,806
Lampung	89,018	78,725	70,959	64,635
Jakarta	103,279	n.a.	76,747	n.a.
West Java	95,017	86,143	71,868	69,287
Central Java	85,667	78,897	66,306	62,559
Yogyakarta	93,078	83,872	70,168	65,805
East Java	85,777	80,496	66,692	64,300
Bali	99,748	94,857	76,004	74,412
West Nusa Tenggara	88,654	85,369	70,746	70,043
East Nusa Tenggara	84,639	78,923	66,198	62,581
West Kalimantan	94,185	88,768	74,734	74,762
Central Kalimantan	96,364	85,670	78,133	75,145
South Kalimantan	86,907	83,294	70,770	69,687
East Kalimantan	96,989	93,340	74,451	75,178
North Sulawesi	87,165	81,905	69,331	67,417
Central Sulawesi	81,527	77,186	64,463	62,604
South Sulawesi	84,734	74,446	66,143	61,867
Southeast Sulawesi	87,269	80,415	67,273	65,338
Maluku	102,522	100,413	76,575	78,545
Papua	88,593	98,102	70,747	74,845

Rp = rupiah
Source: Pradhan et al. 2001.

Appendix 1.3 OLS Regression Results of the Consumption Correlates Model

Predictors	Urban Areas	Rural Areas
Asset Ownership		
this household owns a radio	0.076**	0.059**
	[0.014]	[0.007]
this household owns a television	0.089**	0.070**
	[0.015]	[0.008]
this household owns a refrigerator	0.363**	0.269**
	[0.022]	[0.033]
this household owns jewelry	0.099**	0.071**
	[0.014]	[0.007]
this household owns a satellite dish	0.158**	0.172**
	[0.041]	[0.033]
this household owns a motorcycle	0.221**	0.262**
	[0.021]	[0.015]
this household owns a car	1.342**	0.722**
	[0.058]	[0.082]
Animal Ownership		
this household owns chicken	-0.077**	0.024**
	[0.016]	[0.008]
House Characteristics		
roof of the house is made from tile	0.102**	
	[0.023]	
wall of the house is made from concrete	0.157**	0.061**
	[0.014]	[0.009]
floor of the house is dirt floor		-0.054**
		[0.008]
this household's source of water is from protected well or water pump	0.078**	0.045**
	[0.015]	[0.009]
toilet type of the house is flush	0.093**	0.084**
	[0.014]	[0.011]
this household uses its own toilet	0.094**	0.031**
	[0.015]	[0.007]
this household has electricity		0.092**
		[0.008]
Household Characteristics		
household head age		0.015**
		[0.002]
household head age squared		-0.000**
		[0.000]
spouse age	-0.016**	
	[0.002]	
spouse age squared	0.000**	
	[0.000]	
household head finishes primary education	0.168**	0.030**
	[0.017]	[0.008]
household head finishes junior secondary education	0.245**	0.092**
	[0.022]	[0.019]
household head finishes senior secondary education	0.395**	0.150**
	[0.026]	[0.019]
household head finishes tertiary education	0.734**	0.292**
	[0.046]	[0.042]
spouse finishes primary education	-0.123**	-0.038**
	[0.021]	[0.009]
spouse finishes junior secondary education	-0.178**	-0.051**
	[0.029]	[0.018]
spouse finishes senior secondary education	-0.214**	
	[0.033]	
at least one school-age child (6–15 years old) in this household has dropped out of school		-0.022**
		[0.008]

(continued on next page)

Appendix 1.3 continued

Predictors	Urban Areas	Rural Areas
household size	-0.605**	-0.378**
	[0.020]	[0.009]
household size squared	0.036**	0.023**
	[0.002]	[0.001]
dependency ratio of this household is more than 0.5	-0.068**	-0.058**
	[0.024]	[0.008]
spouse is working	0.072**	
	[0.016]	
at least one school-age child (6–15 years old) in this household is working	0.170**	0.057**
	[0.046]	[0.011]
household head is working in the formal sector		0.053**
		[0.011]
every household member has different clothing for different activities	0.168**	0.144**
	[0.028]	[0.012]
when a member in this household is sick, s/he is treated with modern medicine		0.048**
		[0.010]
Consumption Pattern		
every household member eats at least twice a day	0.176**	
	[0.053]	
this household consumed beef in the past week	0.348**	0.232**
	[0.031]	[0.024]
this household consumed egg in the past week	0.078**	0.111**
	[0.015]	[0.008]
this household consumed milk in the past week	0.405**	0.353**
	[0.022]	[0.023]
this household consumed biscuit in the past week	0.155**	0.064**
	[0.026]	[0.013]
this household consumed bread in the past week	0.128**	0.069**
	[0.018]	[0.010]
this household consumed banana in the past week	0.120**	0.114**
	[0.024]	[0.012]
this household consumed *tiwul* in the past week		-0.052**
		[0.018]
Constant	2.987**	1.335**
	[0.070]	[0.043]
Province dummy variables included	Yes	Yes
Number of observations	23,847	34,649
R-squared	0.44	0.36

** Significant at 1%
[ ] Robust standard errors in brackets
Note: Dependent variable real per capita expenditure is transformed into logarithmic value.
Source: Authors' calculation based on 2002 SUSENAS.

CHAPTER 2

Poverty Predictor Modeling in Indonesia: A Validation Survey

Bayu Krisnamurthi, Arman Dellis, Lusi Fausia, Yoyoh Indaryanti, Anna Fatchia, and Dewi Setyawati

Introduction

The objective of this chapter was to assess and verify the explanatory or predictor variables used for determining the poor. The predictor variables were based on the earlier results of the poverty predictor modeling (PPM) exercise using Indonesia's National Socioeconomic Survey (SUSENAS) discussed in Chapter 1 of this book. The PPM results were used as the basis of the analysis. The verification process was done using a local assessment and survey. The overall results were then analyzed for their significance in determining poverty, especially their usefulness in identifying the poor and improving poverty targeting.

Data and Approaches

Data used in this study emanated from a 2005 sample survey[1] of households in Bogor, West Java, and Tangerang, Banten. The sample included 624 households selected from two groups, i.e., households which were covered in the SUSENAS and households which were not covered in the SUSENAS. For comparison, the secondary data of SUSENAS 2004 for the two districts selected were used as the benchmark for classifying the households into poor and nonpoor.

The poverty predictor variables examined in this study were classified according to the following characteristics:

- ownership of electronic equipment (radio, TV, etc.);
- level of education;
- consumption pattern (no consumption of milk, meat, biscuits, or bread in a week, do not get two meals a day);
- household dependency ratio of more than 0.5;

1 The questionnaire used in the pilot survey can be downloaded at http://www.adb.org/Statistics/reta_6073.asp.

- household attributes (earth floor, impermanent walls, no sanitary facilities, no electricity, etc.);
- main source of income coming from informal sectors; and,
- level of health (cleanliness of clothing, medication).

These variables are similar to those used in the three methods discussed in the previous chapter which were found to be significant in explaining poverty.

In addition, as a complementary measure for deducing information about household poverty status, independent assessments based on four local sources were also used to better view and assess poverty. The perceptions about household poverty status are taken from respondents, respondents' neighbors, local authorities, and enumerators.

The respondent could be one of the most reliable sources of information in assessing whether he or she is poor or nonpoor. Neighbors are another source of information that are considered to be very reliable in judging a respondent's poverty status. The local authorities, as the bureaucracy closest to the respondent, are also an important source of information in this aspect.[2] Lastly, the assessment of the enumerators, who visit the households during the survey, is also important as they are an objective source of information. These assessments, to some extent, can be used for comparison. Among all these factors, the perception of the household respondent is considered most reliable and is given a greater weight (2) than the perceptions of the other three sources which are each given a weight of 1. Setting greater weight to the respondent's perception is deliberate; it aims to improve certainty in determining the poverty status of the respondent.

With this weighting system, the lowest poverty score would be 0, which means that all sources of information perceive that the respondent household is nonpoor. In contrast, the greatest score would be 5 if all sources perceive that the respondent household is poor. If the sum of the weights of perceived poverty is 3 or more, the household is classified as poor. The result of the weighting process for all respondents is presented in Table 2.1.

Using the perception method, 363 of the total 624 household samples were classified poor and 261 nonpoor–with all four sources mostly agreeing on the classification of the households as poor or nonpoor. For example, as many as 251 of the 363 poor households were assigned a local perception weight of 5, which implies that all the sources consider these households as

2 However, uncertainty may arise due to, for instance, the presence of conflicts of interest, which tend to distort the assessment of whether the respondent is really poor.

Table 2.1 **Assessing Poverty by Using the Weighted Perception Method**

Poverty Assessment from Local Perception	*Sum of the Weight of Perceived Poverty*	*Areas*		
		Rural	*Urban*	*Rural+Urban*
Nonpoor	0	70	86	156
	1	21	14	35
	2	33	37	70
Total		124	137	261
Poor	3	38	31	69
	4	24	19	43
	5	126	125	251
Total		188	175	363
Total Respondents		312	312	624

Source: Authors' calculation.

poor. Similarly, 156 of the 261 nonpoor households were classified as such by all the sources. While perception studies are regarded as subjective by many analysts, the consensus on the poverty status of the majority of households by all sources is noteworthy and points to the usefulness of such studies.

Data Analysis Method

Data collected from the field survey were analyzed through quantitative and qualitative methods to validate variables that could be used as predictors. The quantitative method is based on the application of the poverty line based on the household's expenditures and the qualitative method is based on the perceptions of the local people in identifying the poor.

Qantitative Approach

The identification of poverty predictor variables is done by using a logistic (logit) regression model with the household poverty status of poor and nonpoor as the dependent variable (see also the discussion on Method 2 in Chapter 1 of this book). The difference between logistic and probit is that logistic analysis is based on log odds while probit uses cumulative normal probability distribution. The logistic model can be derived from the logistic probability function or opportunity spread function.[3] The probability of a respondent being poor or nonpoor can be formulated as:

$$\pi_i = \frac{e^{g(x)}}{1+e^{g(x)}}$$

$$= \frac{1}{1+e^{-g(x)}}$$

3 Logistic regression calculates changes in the log odds of the dependent variable and not changes in the dependent variable itself as in ordinary least squares regression.

Where

π_i = likelihood of a respondent having the status of poor.

$g(x) = a + bX$

indicates how quickly the probability changes with changing a single unit of X. Because the relation between X and π_i is nonlinear, the parameter b does not have a straightforward interpretation as it does in the ordinary linear regression.[4]

By taking the natural logarithm from the ratio between the probability of a respondent having the status of poor and that of nonpoor, it then follows that:

$$\ln \frac{\pi_i}{1-\pi_i} = g(x)$$

Such an equation can be determined using the maximum likelihood estimation technique specific for the logistic model which is provided in several statistics and econometrics computer programs such as Microfit (Pesaran and Pesaran 1997).

To meet the logit model requirement, the poverty status assessment results using the weighting system must be recategorized into two categories (binary scale), i.e., poor and nonpoor. Nonpoor respondents are those who have scores of 0–2, while poor respondents are those with scores of 3–5. To classify them as binary-scale variables, the nonpoor respondent is assigned the score of 0, and the poor respondent is given the score of 1. Once this is done, the estimation for validation purposes can then be conducted.

The estimation of the logit model is divided into two, for two respondent groups:

- the logit model for all respondents whose poverty status appraisal was based solely on the perception of the local community and enumerator, and
- the logit model for respondents whose poverty status appraisals are consistent between the local community's perception and the poverty-line assessment based on household expenditures.

Logit model estimations for both groups are then further defined by location: rural, urban, and total. Such divisions are made to identify the

4 See http://luna.cas.usf.edu/~mbrannic/files/regression/Logistic.html.

possibility of a difference of poverty predictors between urban and rural areas. In rural and urban area regression equations, the variable *district* is added as dummy variable; in the combination regression equation, the variable *area* is added as its dummy variable to mean either rural or urban.

Variables used in the validation are the same as those used in the initial stage of PPM. These variables were classified according to:

- ownership of farm animals, which comprise livestock (cattle, buffalo, horses, or pigs), goats, sheep, lambs, poultry (chickens or ducks), and fish;
- ownership of assets such as electronic equipment (radios or tape players, TVs, and satellite dishes), refrigerators, and telephones; vehicles (bicycles, motorcycles, cars or trucks, and carriages); and tools for production (hand tractors, crop machines, pumps, etc.);
- ownership of sanitary facilities (toilets), clean- and potable-water facilities, electrical connections, and cooking facilities;
- physical condition of the house based on floor area, and materials of the floor, walls and roof;
- household characteristics such as age, family size, members with formal education, members who are elementary school dropouts, working members, average educational attainment, dependency ratio, and occupation of the head of the family (formal or informal); and
- consumption pattern for food and nonfood items or characteristic such as rice, meat, eggs, and fish per week; clothes bought in a year; incidence of illness among members in the past six months or the previous year; and the practice of seeking medication when ill.

For each regression, a stepwise procedure is used to minimize the number of variables included in the model. Tests on reliability in predicting poverty status are also done by using cross tabulation between the predicted poverty status as a result of logit model and the status based on the local perception.

Qualitative Approach

The qualitative approach is performed to explain the various characteristics of the respondents, which comprise ownership of livestock, poultry, fish, and assets; physical condition of the house and facilities; household characteristics; and food consumption, health, and nutrition. Qualitative analysis is implemented using cross tabulation between respondents' poverty status, various characteristics, and respondents' perception.

Results

Poverty Classification and Verification

Poverty verification in this study is based on two assessment approaches: local perception and household expenditure using predetermined poverty indicators. For each approach, classifying the household respondents into poor and nonpoor is attempted.

Poverty Verification Based on Local Perception. Table 2.2 shows that based on local perception, 58.2 percent of household respondents are considered poor. Of this number, 30.1 percent were perceived to be in rural areas while 28.1 percent were in urban areas. Corollary to this, the perception is that there are more nonpoor households in the urban areas (22.0 percent) than in the rural areas (19.9 percent).

Table 2.2 **Classifying Poor and Nonpoor Households by Using the Local Perception Approach**

Respondent Status	*Area*		
	Rural	*Urban*	*Rural+Urban*
Poor	188	175	363
	30.1 %	28.0 %	58.2 %
Nonpoor	124	137	261
	19.9 %	22.0 %	41.8 %
Total	312	312	624
	50.0 %	50.0 %	100.0 %

Source: Authors' calculation.

Poverty Verification Based on Household Expenditures. Recalculating the actual poverty line is considered necessary because of the dynamic nature of the conditions of poverty. It is acknowledged that, after a year, the condition of a household may change as a result of a change in the household's expenditures. Taking this into account, the verification of the SUSENAS data for 2004 is also based on the expenditures of the household.

Poverty verification based on household expenditures is measured by taking the average threshold of monthly household expenditure per capita, which is Rp130,927[5] for Bogor and Rp132,108 for Tangerang in 2004. This implies that households with per capita expenditures lower than the thresholds for each of these districts will be considered poor, thus, these thresholds are in effect pseudo poverty lines.

The results of poverty verification based on household expenditures as shown in Table 2.3 indicate that 58.7 percent of household respondents are poor, and 41.3 percent are nonpoor. Furthermore, the number of poor households in rural areas (36.2 percent) is higher than in urban areas (22.4 percent) and the number of nonpoor households in rural areas (13.8 percent) is less than in urban areas (27.6 percent).

5 Rp stands for rupiah; US$1 is roughly about Rp9,000 (2004).

Poverty Verification Based on Both Assessment Approaches. The consistency, or the lack of it, of the poverty verification results based on local perception and household expenditures can be tracked when the results are presented in a single matrix. A cross tabulation of the results from the two different assessment methods is thus presented in such a matrix in Table 2.4. The table shows that based on local perception and household expenditure assessments, 43.1 percent of the households in rural and urban areas combined are poor and 26.3 percent are nonpoor. The rest of the observations show inconsistent results between the two assessment approaches. About 15.1 percent of the households are poor based on local perception, but they are considered nonpoor based on expenditure. On the other hand, 15.5 percent of the households are perceived as nonpoor by the local community, but, based on expenditure, they are considered poor. It is clear from these observations that results using expenditure data to identify the poor will differ by about 15.0 percentage points compared with the result using local perception, and vice versa.

Table 2.3 **Classifying Poor and Nonpoor Households by Using the Expenditure Approach of the Pilot Survey**

Respondent Status	*Area*		
	Rural	*Urban*	*Rural+Urban*
Poor	226	140	366
	36.2%	22.4%	58.7%
Nonpoor	86	172	258
	13.8%	27.6%	41.3%
Total	312	312	624
	50.0%	50.0%	100.0%

Source: Authors' calculation.

Table 2.4 further reveals that verification results of SUSENAS data for 2003/04 are consistent in the estimation of the proportion of poor based on pilot survey. Verification results based on local perception show the 58.2 percent of the respondents are actually poor and 41.8 percent are nonpoor. While verification based on recalculating household expenditures (using the pseudo poverty line) has fairly similar results: 58.7 percent of the households are poor and 41.3 percent are nonpoor.

Table 2.4 **Classifying Poor and Nonpoor Households by Using the Local Perception and Household Expenditure of the Pilot Survey Approaches**

		Household Expenditures		
		Poor	*Nonpoor*	*Total*
Local Perception	*Poor*	269	94	363
		43.1%	15.1%	58.2%
	Nonpoor	97	164	261
		15.5%	26.3%	41.8%
	Total	366	258	624
		58.7%	41.3%	100.0%

Source: Authors' calculation.

Poverty Estimation. The results of poverty estimation in rural and urban areas are, interestingly, consistent with the verification of SUSENAS data for 2004 and in the assessment approaches based on local perception and household expenditures. Even though there are slight differences, the three assessment methods are in general relatively consistent, as seen in Table 2.5.

Verification using the 2004 data shows that 48.7 percent of households (25.8 percent in rural and 22.9 percent in urban areas) are classified as

Table 2.5 **Classifying Poor and Nonpoor Households by Using SUSENAS Data, Local Perception, and Household Expenditures of the Pilot Survey Approaches**

Area	SUSENAS			Household Expenditures			Local Perceptions		
	Poor	Nonpoor	Total	Poor	Nonpoor	Total	Poor	Nonpoor	Total
Rural	25.8	24.2	50.0	36.2	13.8	50.0	30.1	19.9	50.0
Urban	22.9	27.1	50.0	22.4	27.6	50.0	28.0	22.0	50.0
Rural+Urban	48.7	51.3	100.0	58.7	41.3	100.0	58.2	41.8	100.0

SUSENAS = National Socioeconomic Survey
Source: Authors' calculation.

poor (with low-expenditure households as a proxy for poverty). However, the results are slightly different if the verification is conducted using results of recalculations based on household expenditures or local perception. About 58.7 percent households are considered poor based on expenditure assessment, i.e., 36.2 percent in rural and 22.4 percent in urban areas. The results from using local perception verification have similar results: 58.2 percent of households are considered poor, i.e., 30.1 percent in rural and 28.0 percent in urban areas.

The above information also confirms the dynamic aspect of poverty. There is a difference of about 10 percentage points between the results of the verification from pilot survey using the data and the recalculation of the poverty line based on household expenditures. About 48.7 percent households are poor according to the SUSENAS data, but 58.7 percent are poor according to the assessment based on expenditure. This means that in one year, i.e., from the 2002 SUSENAS to the 2004 SUSENAS, about 10 percent of households experienced a fall in their total expenditures and became poor. This highlights the vulnerability of people who are above but close to the poverty line.

When the SUSENAS data is verified using the results of local-perception assessment, there is a slight difference in the ratio of poor and nonpoor household groups. Based on the 2004 data, about 48.7 percent of households are poor; but, based on local perception, 58.2 percent households are considered poor. This means that 10 percent of the households considered nonpoor in the 2004 are perceived as poor by the local communities.

Predictability of Poverty Variables

Estimation Results of the Local Perception Logit Model. The results of a logistic regression model of respondents' poverty status based only on local perception (Appendix 2.1) show that the logistic models for rural, urban, and total respondents have a relatively small pseudo R-squared value. The retained predictors only explain 44.1 percent of the respondents' poverty status in rural areas and 52.3 percent in urban areas. The combination of rural and urban respondents resulted in an even smaller pseudo R-squared value (38.1 percent). Small R-squared values are, however, usually found in regression models with dichotomous variables. In predicting power, the result shows 83.3 percent is true for the model for rural areas, 86.5 percent for urban areas and 79.5 percent for the total. The following is a summary on the predictability of the retained variables.

Asset Ownership. The variables for ownership of refrigerators, TVs, and motorcycles have positive values and are significant for rural areas, while the ownership of TVs and motorcycles are significant for the urban areas. The regression for total respondents shows that the three asset-ownership variables are also significant and consistent. Since the variables are specified in terms of nonpossession of these assets, the positive values mean that households which do not have refrigerators, TVs, and motorbikes have a higher probability of being poor compared with those who have these assets.

House Characteristics. House characteristics in rural and urban areas are very different. In rural areas, the type of wall in a house has positive values, meaning that if a house does not have a brick concrete wall the household is more likely to be poor. In urban areas, the significant variable is floor area. The more spacious the house, the less likely the household is poor.

House Facility. Toilet ownership is significant in the three models and has positive values. This implies that the poor are less likely to have a toilet and nonpoor households tend to have their own toilet.

Household Characteristics. The retained variables for the model for rural areas are: a family member dropped out from elementary school, the head of family works in the informal sector, and the household dependency ratio is no more than 0.5. The first variable has a positive effect on rural poverty. The last two variables are significant in equations for both rural and urban areas as well as for total respondents. On the other hand, variables that are significant and have positive values in urban areas are: having household members who did not complete their primary education and the square of the number of working household members. A household's size has a significant and positive effect on poverty, while the number of household members with schooling has a negative effect for rural and urban areas combined. Therefore, poor households are identified as having many family members, a member or members who have dropped out of primary school, a relatively small number of working household members or a high dependency ratio, and a main wage earner who is working in the informal sector.

Consumption, Food, Nutrition, and Health. In the last group of variables, having insufficient rice (staple food) and not having eaten meat, eggs, and fish in the reference period are a positive and significant poverty predictor variable in all areas. The use of medical facilities and paramedics is also a significant poverty predictor variable with a positive coefficient in rural and urban areas combined.

Characteristics of Location. The location characteristic is a significant dummy variable. Findings shows that a rural community in Bogor has a lower probability of being poor than a rural community in Tangerang. On the other hand, an urban community in Bogor has a higher probability to be classified as poor than an urban community in Tangerang. The difference could be related to the characteristics of the two districts. Bogor is basically agrarian, with ample employment opportunities in the rural area. Tangerang, on the other hand, is basically industrial, with better employment opportunities in urban areas. This finding highlights the importance of taking characteristics of region and location into account in developing the poverty predictor model.

Estimation Results of the Perception-Expenditure Logit Model. The perception-expenditure logit model refers to the logit model estimation for respondents whose poverty status based on their expenditure is consistent with the local community's perception. The results (Appendix 2.2) are similar to the results from the poverty estimation model in terms of variable and estimation procedures.

Analyzing respondents with consistent perception-expenditure results from the model, shows that the pseudo R-square value increased compared with the previous estimate of 38.1 percent. In rural areas, the model can be used to explain 66.4 percent of the respondents' poverty status; in urban areas, 76.6 percent can be explained; and, for all respondents, 66.3 percent can be explained. In addition, there are some new predictor variables that resulted from this model. The variables of ownership of cows in rural areas and sheep in urban areas were found to be significant in predicting poverty.

The variables of TV and motorbike ownership remain significant in rural areas. In urban areas, however, the ownership of telephones, radios or tape recorders, and motorbikes are significant. For total respondents, however, the ownership of a radio or tape recorder becomes insignificant.

House ownership was not significant among rural, urban, or total respondents and so it was not used as a poverty predictor variable in the perception-expenditure model. On the other hand, the use of simple cooking utensils powered by wood is a poverty indicator in rural areas. In urban areas, the ownership of toilet is a significant predictor variable, which is consistent with the finding from the poverty estimation discussed in the previous section

Household-specific variables show that family size, education level of household members, and household-head employment are important poverty predictor variables. Having rice and eating meat, eggs, and fish in the past week are consistent with the previous estimation result. A new variable on health appears in urban areas: a household whose members are frequently sick has a higher probability of being poor.

In general, the estimate for the perception-expenditure model results in some main poverty predictors such as:

- non-ownership of electronics (TV, radio, or tape recorder), refrigerator, telephone, or motorbike;
- house has no personal toilet and the household uses simple cooking utensils fired by wood in rural areas;
- large family size, small number of household members in school, and low average education level of household members;
- family earner works in the informal sector and relatively small number of working household members (high dependency ratio, less than 0.5) and;
- not owning sufficient staple food (rice), nutrition deficiency (unable to consume meat, eggs, and fish at least once a week), and poor health and inability to visit a general practitioner or hospital for medical care.

Compared with the SMERU result based on the SUSENAS data, several variables out of the seven indicators of poverty are consistent except household characteristics. In this study, family size is an important poverty indicator compared with the SMERU result. In addition, household's inability to have sufficient rice and use of firewood as a fuel are also poverty predictors in rural areas in this study but not in SMERU.

Accuracy of the Predictor Variables. The capability of predictor variables to explain poverty can also be seen by comparing the actual poverty status of the household with the predicted poverty status. The predictive value for the dependent variable is distributed as *0* or *1*, thus, requiring households to be classified as poor or nonpoor. This means a clustering process can be done automatically using the Microfit computer program. In this context, households with more than 50 percent probability of being poor are classified as poor and, conversely, nonpoor if the probability is less than 50 percent.[6]

By cross tabulating the actual and predicted household poverty status, two sets of results can be obtained. The first is shown on Table 2.6 based on

6 This classification technique is commonly applied in econometrics (Verbeek 2000). The classification used here is slightly different than the classification used in the study by Sumarto, Suryadarma, and Suryahadi (Chapter 1 of this book). In that study, households with more than 50 percent poverty probability were classified as poor (see also Sumarto 2004).

local community's perception and the second is shown in Table 2.7 based on consistent perception-expenditure respondents.

Table 2.6 indicates that 47.8 percent of total households in rural and urban areas together are classified as poor and 29.5 percent as nonpoor. The accuracy and effectiveness of poverty indicators can be obtained by adding the primary diagonal elements in the table. For example, the effectiveness of the poverty indicator[7] for rural areas is 83.4 percent–the sum of the percentage of households that were predicted to be nonpoor and were actually nonpoor (29.2 percent) and the percentage that were predicted to be poor that were actually poor (54.2 percent). For urban and total respondents, therefore, the effectiveness of the poverty indicator is 86.6 percent, and 77.3 percent, respectively. The numbers demonstrate the combined accuracy of predicting the poor and nonpoor. Note that 9.9 percent and 7.4 percent of households, who are actually nonpoor, were predicted to be poor in rural and urban areas, respectively. On the other hand, 6.7 percent and 6.1 percent of households who are actually poor, were predicted as nonpoor in rural and urban areas, respectively.

Table 2.6 **Predicting Poor and Nonpoor Using the Logit Model for All Respondents**

		Predicted					
		Rural		*Urban*		*Rural + Urban*	
		Nonpoor	*Poor*	*Nonpoor*	*Poor*	*Nonpoor*	*Poor*
Actual	*Nonpoor*	29.2%	9.9%	36.9%	7.4%	29.5%	12.3%
	Poor	6.7%	54.2%	6.1%	49.7%	10.4%	47.8%

Source: Authors' calculation.

Table 2.7 **Predicting Poor and Nonpoor Using the Logit Model for Respondent with Consistent Poverty Status Based on Perception-Expenditure Approaches**

		Predicted					
		Rural		*Urban*		*Rural+Urban*	
		Nonpoor	*Poor*	*Nonpoor*	*Poor*	*Nonpoor*	*Poor*
Actual	*Nonpoor*	20.3%	4.5%	35.9%	13.4%	32.3%	5.5%
	Poor	2.5%	72.8%	4.3%	46.3%	3.5%	58.7%

Source: Authors' calculation.

In the group of respondents having consistent poverty status based on perception and expenditure, the effectiveness of prediction is higher, i.e., 93.1 percent, 82.2 percent, and 91.0 percent for rural, urban, and total respondents, respectively. As a result, the prediction margin of error is minimized at 7 percent for rural and total households, and 17.8 percent for urban households. Based on this result, the effectiveness of significant variables in the logit model is quite high and could be used as poverty predictors in rural and urban areas.

7 This refers to the sum of the primary diagonal elements in Table 2.6.

Appendix

Appendix 2.1 **Results of Logit Model Using SUSENAS Data**
(Dependent Variable: 1 = Poor, 0 = Otherwise)

Predictor	Rural	Urban	Rural-Urban
Asset Ownership			
household has no refrigerator (1 = yes, 0 = otherwise)	2.5497 * (2.7777)	-	0.99917 ** (2.3669)
household has no television (1 = yes, 0 = otherwise)	.94076* (2.7540)	1.2358* (2.9711)	0.75323* (3.1516)
household has no motorcycle (1 = yes, 0 = otherwise)	1.7534* (3.5333)	1.2285** (2.2257)	1.3661 * (4.1772)
House Characteristics			
area of the floor of the house (in m2)	-	-0.0081** (-2.0726)	-
wall of the house is not made from concrete brick (1 = yes, 0 = otherwise)	1.4996* (4.2669)	-	0.63639 * (2.8749)
House Facility			
household has no toilet (1 = yes, 0 = otherwise)	0.78152 ** (2.0539)	1.4393* (3.6155)	1.0624* (4.4039)
Household Characteristics			
Household size (in person)	-	-	0.23871* (3.0599)
household members schooling (in person)	-	-	-0.26253*** (-1.9314)
average household education did not finish primary school (1 = yes, 0 = otherwise)	-	1.2100* (2.8863)	1.0800* (4.6711)
household members have dropped out of primary school (1 = yes, 0 = otherwise)	0.91053 ** (2.1784)	-	-
square of number of household members who are working (in person)	-	0.18311* (2.9057)	-
head of household work in informal sector (1 = yes, 0 = otherwise)	2.1656* (4.7848)	1.6854* (3.5813)	0.67244** (2.0749)
dependency ratio of this household is less than 0.5 (1 = yes, 0 = otherwise)	0.9246** (2.1262)	1.9828* (3.9781)	0.90756* (3.3196)
Consumption, Food, Nutrition and Health			
this household has insufficient rice consumption (1 = yes, 0 = otherwise)	2.2314** (2.5507)	0.89972 (1.5858)	1.6790* (4.0677)
household that has not consumed meat, egg or fish in the past week (1 = yes, 0 = otherwise)	2.3752* (4.3885)	1.5896* (3.1905)	0.72304** (2.4352)
treated at the local health centre (Puskesmas). medical aide (mantri), midwife (bidan) or traditionally (1 = yes, 0 = otherwise)	-	0.72577*** (1.8511)	-
Dummy Variable for District and Rural-Urban Area			
dummy variable for district (1 = Bogor, 0 = otherwise)	-1.4041* (-3.5623)	2.1659* (4.4066)	-
dummy variable for rural-urban area (1 = rural, 0 = otherwise)	-	-	-0.52526 (-2.2028)
Constant	-6.6374* (-5.6238)	-6.4282* (-6.6906)	-5.1900* (-8.3197)
Goodness of fit	0.83333	0.86538	0.79487
Pseudo R-squared	0.44112	0.52338	0.38120
Numbers of Observation	312	312	624

*** Significant at 10%; ** Significant at 5%; * Significant at 1%
SUSENAS = National Socioeconomic Survey
Source: Authors' calculation based on 2004 SUSENAS.

Appendix 2.2 **Logit Model Results with Consistent Poverty Status Based on Perception and Expenditure Approaches (Dependent Variable: 1 = Poor, 0 = Otherwise)**

Variable	Rural	Urban	Rural-Urban
Animal Ownership			
household has no goat (1 = yes, 0 = otherwise)	-	1.9877** (2.2427)	-
household has no cow or buffalo (1 = yes, 0 = otherwise)	2.6187** (2.3838)	-	-
Asset Ownership			
household has no telephone (1 = yes, 0 = otherwise)	-	5.8899* (3.3749)	3.1160* (2.6862)
household has no radio and tape recorder (1 = yes, 0 = otherwise)	-	1.8490* (2.9378)	-
household has no refrigerator (1 = yes, 0 = otherwise)	-	-	2.4053* (2.8421)
household has no television (1 = yes, 0 = otherwise)	1.7068 ** (2.2640)	-	.84419 ** (2.0015)
household has no motorcycle (1 = yes, 0 = otherwise)	2.3037** (2.1901)	5.2100* (3.1299)	2.1997 * (3.4043)
House Facility			
household uses firewood (1 = yes, 0 = otherwise)	2.6151* (3.5262)	-	-
Household has no toilet (1 = yes, 0 = otherwise)	-	2.4252* (3.1952)	0.95967** (2.4583)
Household Characteristics			
household representative age (in year)	-	-	0.0249*** (1.9341)
household size (in person)	1.2020* (3.6570)	1.1673* (4.5025)	0.86228* (5.1340)
household members at school (in person)	-1.1316** (-2.3962)	-	-0.58246** (-2.1169)
average household education not graduating primary school (1 = yes, 0 = otherwise)	1.6499** (2.4445)	-	0.72488*** (1.8308)
head of family has worked in informal sector (1 = yes, 0 = otherwise)	3.2554* (3.0022)	6.2795* (4.4332)	2.8647* (4.4632)
Dependency ratio of this household is less than 0.5 (1 = yes, 0 = otherwise)	-	-	0.86421*** (1.8269)
Consumption, Food, Nutrition and Health			
household insufficient rice consumption (1 = yes, 0 = otherwise)	3.3702** (2.2405)	-	2.0157* (2.6448)
household has not consumed meat, egg or fish in the past week (1 = yes, 0 = otherwise)	1.6757** (1.9750)	3.6518* (3.4965)	1.6350* (2.6765)
household member sick in the past year (1 = yes, 0 = otherwise)	-	2.2932* (2.9120)	.81583*** (1.8044)
treated at village clinic, medical aide (mantri), nurse or traditionally (1 = yes, 0 = otherwise)	-	.	0.96881** (2.1529)
Dummy Variable for Regency			
dummy variable for regency (1 = Bogor, 0 = otherwise)	-4.2598* (-3.7720)	0.5729* (2.8348)	-
Constant	-10.7518* -4.3221)	-27.7208* (-5.1578)	-15.9654* (-6.9889)
Goodness of fit	0.93069	0.93506	.90993
Pseudo R-squared	0.66390	0.75600	.66315
Numbers of Observation	202	231	433

*** Significant at 10%; ** Significant at 5%; * Significant at 1%
Source: Authors' calculation.

CHAPTER 3

Identifying Poverty Predictors Using China's Rural Poverty Monitoring Survey

Sangui Wang, Pingping Wang, and Heng Wang

Introduction

As the world's largest developing country, the People's Republic of China (PRC) has a large rural poor population. Using the official poverty line and household income data, the number of rural poor people was estimated at 19 million by the end of 2005. Using a higher poverty line (close to the $1-a-day standard), the number of poor is estimated to be 82 million (KI 2007). Estimation based on household consumption expenditure leads to a much higher number of rural poor (Wang, Li, and Ranshun 2004).

Though rural poverty reduction has been dramatic because of continuing economic growth and targeted poverty reduction interventions sponsored by different government institutions in the past two decades, major challenges exist in identifying the poor for more effective poverty intervention schemes. Because there is no reliable household-level information in terms of income and expenditure available for local areas, the PRC has long been relying on geographic targeting (at county and village levels) for its poverty reduction programs. This has led to severe undercoverage and leakage problems in program and project implementation (Sangui 2005). Alternative ways to easily identify individual poor households for more effective poverty targeting are urgently needed in the PRC.

Poverty predictor modeling (PPM), established by using household survey data and modern econometric analysis, is one alternative that can be applied to individual poverty targeting (Ward, Owens, and Kahyrara 2002). This chapter discusses the methods and processes of PPM for the PRC. The main purpose of this modeling exercise was to estimate the correlates of poverty at the household level. For practical reasons, poverty predictor variables included–and eventually found significant in the modeling exercise–were non-income and other expenditure indicators that are easily collected.

Data and Methods

Data

In this study, the data set from the 2002 China Rural Poverty Monitoring Survey (CRPMS) collected annually by the Rural Survey Organization (RSO) of the National Bureau of Statistics was used to establish the poverty predictors. CRPMS is conducted in rural areas, hence, data can better reflect the living conditions and household characteristics of the poor than other existing but inaccessible data sets in the country. In addition, survey results provide more program- and policy-relevant information needed in the modeling.

The questionnaire used in the CRPMS is similar to the one used in the Rural Household Survey, which has been the source of official poverty statistics in rural PRC. It includes detailed household and individual information on income and expenditures, household demographics, production, assets, education, and employment. Additional information on rural infrastructure and poverty programs are also collected at the village and household levels. The data collected from CRPMS have mainly, since 2000, been used by RSO to produce an annual Rural Poverty Monitoring Report.

The 2002 CRPMS has a large sample size of 50,000 households. Excluding the households with missing values, the total sample would be 45,960 households. For comparison and robustness tests of the regression models, the sample was split into two subsamples: Data1 and Data2. Village codes were randomly assigned to the sample villages and the splitting of the sample was done by assigning those with odd village codes to Data1 and those with even village codes to Data2. Through the existing sampling design, each poor county with 5–10 sample villages and 10 households in each village are randomly sampled for the survey. Since the village codes are randomly assigned to the sample villages, the splitting of sample households can be considered a random process.

After splitting the codes, Data1 had 22,845 sample households and Data2 had 23,115 sample households. Their mean per capita consumption expenditures were CNY1,414.76[1] and CNY1,423.69, respectively. The process of identifying the best model was applied to both data sets.

Methods Adopted

Two types of econometric models were used for this PPM effort. The first one was the most commonly used multiple regression model that examines

1 CNY stands for yuan.

the relationship between household expenditure and poverty based on individual, household, and community characteristics. The result identified specific variables (predictors) that were significantly correlated with household living–standard variables (i.e., consumption expenditure or income). The second one was a logistic regression model that predicted the probability of a household being poor or not.

The multiple linear regression models took the form of:

$$y_i = \alpha + \sum \beta_k x_{ki} + e_i$$

Where:

y_i - the dependent variable

x_{ki} - independent variables/predictors

α - the model intercept

β_k - regression coefficients

e_i - random errors

Logistic regression models took the form of:

$$\ell n(\frac{p_i}{1-p_i}) = \alpha + \sum_{k=1}^{k} \beta_k x_{ki}$$

Where:

$p_i = P(y_i = 1 | x_{1i}, x_{2i}, ..., x_{xki})$ is the probability of an event given $x_{1i}, x_{2i}, ..., x_{ki}$.

$\frac{p_i}{1-p_i}$ is the odds of experiencing an event.

As in the PPM for Indonesia (see Chapters 1 and 2 of this book), the regression analysis used a stepwise procedure at the 5-percent level of significance to limit the number of independent variables included in the model. For the multiple regression procedure, a number of diagnostic checks and tests were applied to evaluate the adequacy of the model: normal plots, residual plots, and scatter plots, and the assessment of the variance inflation factor (VIF) for the multicollinearity test. A variable was dropped from the model if the VIF of the variable was greater than 10.

For logistic regression, the goodness-of-fit test was used to check the accuracy of the model. The Hosmer-Lemeshow test (Wang and Zhigang 2001) was also used because the number of covariate patterns was almost the same as the number of observations. This was attributed to a number of continuous independent variables that were employed. The test was carried out by computing the percentile distribution of the predicted probabilities (10 groups based on percentile ranks) and then computing a Pearson chi-square that compares the predicted to the observed frequencies (in a 2 X 10 table). Lower values (and nonsignificance) indicate a good fit of the model to the data.

To examine predictability of the method, sensitivity and specificity (accuracy) tests and graph sensitivity and specificity versus probability cutoffs for identifying the best cutoff points were also used for the two methods.

Identification of Variables

In search of candidate independent variables (predictors) from more than 500 indicators collected by RSO, the empirical study focused on variables which are theoretically and empirically correlated with household welfare variables and poverty status, and are easy to collect. Since there was no intention to estimate the determinants (causality) of household welfare or poverty status, the endogeneity of the independent variables was not a concern.

The identified candidate variables were roughly classified into five groups: household demographics, characteristics of household head, assets and natural resources, activities and access to services, and community characteristics. (Candidate variables selected for the estimation are listed in Appendix 3.1.)

Household income and consumption expenditure data were both collected by the RSO in the CRPMS. However, expenditure was considered to be a better measure of both current and long-term welfare and was employed as the dependent variable in the multiple regression model. Because individuals prefer to smoothen the consumption trend over time, expenditure tends to vary less from year to year than income. Another reason for choosing expenditure is that there are negative values of income in the sample, that is, when household production costs exceed revenues. With negative values, logarithmic transformation is impossible.

For logistic regression, the binary dependent variable is anchored to the consumption expenditure data. When the per capita expenditure of a household is below the poverty line, the household is classified as a poor household, and nonpoor if otherwise.

The official rural poverty line in the PRC is used to classify all the sample households into poor and nonpoor. This is estimated by the RSO and used to calculate the poverty headcount ratio every year. There are two poverty lines, an absolute poverty line and a low-income poverty line. The latter is close to the purchasing power parity–adjusted $1-a-day poverty line of the World Bank. The PRC's poverty lines are not adjusted for regional price differences and the lines are uniform for the whole country. In 2002, the low-income poverty line was CNY869 and the absolute poverty line was CNY627.

Transformation of Variables

To decide whether a transformation of the dependent variable (household consumption expenditure per capita) was necessary, a regression procedure was applied to both untransformed and log form per capita expenditure. Accordingly, it was found that the natural logarithm form increased the R-squared and adjusted R-squared.[2] Thus, the log of per capita expenditure was used in this study.

As for the independent variables, three types of transformation were undertaken: natural logarithm, square rooting, and reciprocation. Inspecting the scatter plot of each transformed-type variable against the log per capita expenditure and the resulting adjusted R-squared, some variables were used in transformed form as indicated in Table 3.1. The rest of the variables were left untransformed.

Table 3.1 **Transformation Scheme for Independent Variables to Reduce Measurement Error**

Variables	Transformation
• Housing acreage	Square root
• Amount of grain stored at home per capita	Square root
• Amount of grain stored at home per capita	Square root
• Number of family members staying at home for six months or more	Natural logarithm

Source: Authors' summary based on the modelling development results.

Results

Multiple Regression Models

Table 3.2 shows the summary results of the stepwise regression for Data1 and Data 2. Models for Data1 and Data2 can only explain 46.2 percent and 46.7 percent, respectively, of the variations in per capita consumption

2 Because the dependent variables are not the same, we can not compare the R-squared directly. But we can calculate the comparable R-squared by transforming the Yi and predicted Yi (Y) and using the formula

$$A_j = \sum_{i=1}^{N} \frac{f_i(a_{ji} - a_i)}{s_i}$$

we find that the comparable R-squared of the log-transformed regressions are much higher (around 0.46) than that of the untransformed regressions (around 0.39).

expenditure. This is actually higher than that of the PPM study for Indonesian data but lower than what has been reported for Viet Nam (see details of the results in Appendixes 3.2 and 3.3).

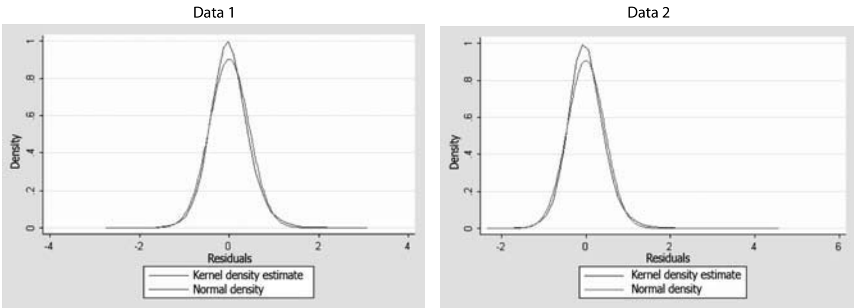
Table 3.2 **Summary Results of Stepwise Ordinary Least Squares Regression for Model Building**

Item	Data1	Data2
Number of observation	22,845	23,315
F-statistics	273.58	282.63
Probability > F	0.0000	0.0000
Adjusted R-squared	0.4621	0.4373

F where the means of multiple normally distributed populations have the same standard deviations.
Note: Data1 and Data2 are subsamples of data used in the model building.
Source: Authors' calculation based on 2002 CRPMS.

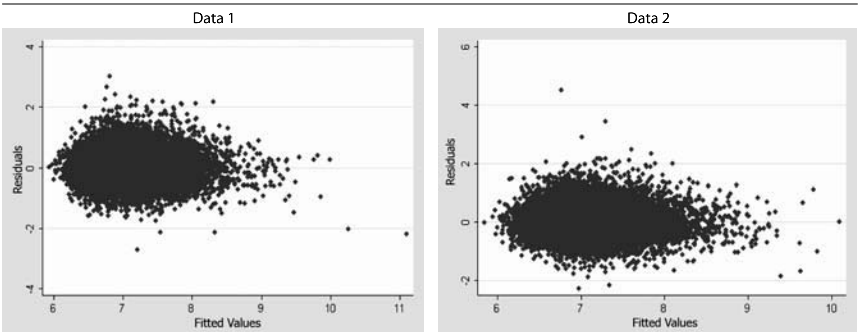
As exhibited in Figure 3.1, distributions of residuals for Data1 and Data2 show that the former is normal while the latter is approximately normal. Next, residual plots in Figure 3.2 reveal that there is no pattern of heteroscedasticity in both Data1 and Data2. This means that on transformation, the assumption of constancy of variance has been satisfied

Figure 3.1 **Normality Plot of Residuals of the Ordinary Least Squares Regression for Data1 and Data2**



Source: Authors' calculation.

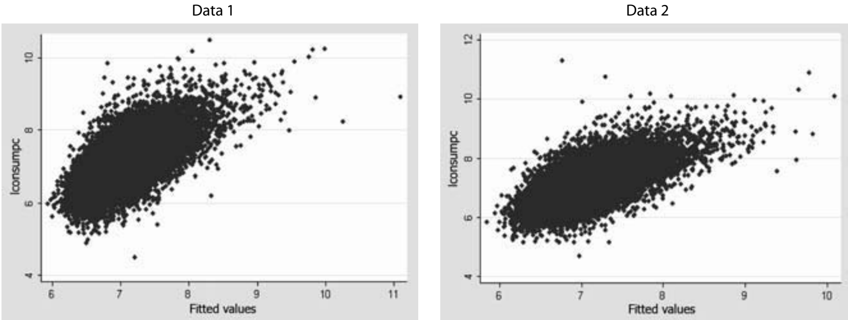
Figure 3.2 **Residual Plot of the Ordinary Least Squares Regression for Data1 and Data2**



Source: Authors' calculation.

by the predicted values of per capita consumption. Figure 3.3 shows that the plotted predicted values as against the actual per capita expenditure not only validated homoscedasticity but also proved nonexistence of outliers

Figure 3.3 **Scatter Plot of Actual Per Capita Consumption Against Predicted Values for Data1 and Data2**



Source: Authors' calculation.

and the independence of the error terms. Results of the VIF (Table 3.3 and 3.4) for the two data sets, revealed that none of the variables generated VIF values greater than 10. Hence, multicollinearity was ruled out and none of the variables were dropped.

Household Demographic Characteristics. This section discusses the results on regression coefficients with an age effect of household members on per capita expenditure. Holding other factors constant, for a household with more members 15–60 years old, the increase in expenditure per capita is higher than a household with more members aged 0–14 years or over 60 years old. Hence, a household with more members aged 15–60 years old is less likely to be poor. This is because individuals of ages 15–60 years are usually more productive than their younger or older counterparts and, hence, can contribute to the household's income pool, which allows household members to consume more.

The composition of households also correlates with the level of expenditure of its members. A household with three generations tends to consume more per member compared with all other kinds of households and is less likely to be poor. In rural PRC, traditional families have three generations under one roof. Not only does this arrangement allow for household savings, but income from rural production of the young and the savings of the old are also shared among the household members.

Also, assuming all other variables stay the same, household consumption per capita is usually higher and the household is less likely to be poor in a household with a larger number of school-age children. A household that can afford to send their children to school is relatively more affluent compared with a comparable household in rural areas where household members have to work on agricultural farms.

Table 3.3 **Variance Inflation Factor of the OLS Regression Using the Data1 Subsample**

Variable	VIF	1/VIF	Variable	VIF	1/VIF
_Ib5_6	7.84	0.12759	_Ipro_43	1.43	0.70040
_Ib5_3	7.07	0.14139	_Ipro_14	1.40	0.71543
_Ib5_4	6.88	0.14538	_Ipro_50	1.39	0.72190
ln_p	5.23	0.19117	c21	1.38	0.72445
_Ib5_2	4.06	0.24601	_Ipro_34	1.37	0.73115
age15_60	4.01	0.24913	b22	1.37	0.73244
age0_14	3.81	0.26217	b19	1.34	0.74477
_Ic13_3	3.79	0.26364	_Ipro_63	1.27	0.78529
b13	3.51	0.28524	a6	1.27	0.78571
_Ipro_65	3.41	0.29307	fuel	1.25	0.79744
b30	3.37	0.29684	b41	1.25	0.80238
_Ic13_2	3.29	0.30366	b26	1.24	0.80784
c7	2.94	0.34025	b21	1.23	0.81521
_Ipro_53	2.48	0.40315	_Ia1_2	1.22	0.81714
_Ib5_7	2.38	0.41949	_Ipro_64	1.20	0.83210
age60	2.29	0.43744	_Ic13_5	1.18	0.84799
_Ic13_4	2.28	0.43893	a57	1.17	0.85573
_Ib5_5	2.06	0.48471	b31	1.17	0.85672
b24	1.97	0.50688	c4	1.16	0.86432
ro_n_b10	1.93	0.51734	b17	1.15	0.86834
studt	1.93	0.51849	leadbus	1.14	0.87359
_Ipro_52	1.87	0.53348	_Ipro_46	1.14	0.87636
b23	1.83	0.54784	a50	1.14	0.87971
a20	1.75	0.57264	b18	1.13	0.88148
spouse	1.68	0.59467	b47pc	1.11	0.89794
a15	1.62	0.61848	b3	1.10	0.90509
b20	1.61	0.62231	_Ipro_22	1.10	0.90640
c5	1.59	0.62851	b7	1.10	0.91096
_Ipro_45	1.58	0.63247	b8	1.08	0.92897
_Ipro_42	1.53	0.65362	b45pc	1.07	0.93294
landpc	1.52	0.65961	b34	1.07	0.93350
_Ipro_41	1.49	0.67194	cashr	1.07	0.93470
b15	1.48	0.67449	bigevent	1.04	0.96371
ro_n_b73	1.45	0.68817	b25	1.03	0.96814
_Ipro_36	1.44	0.69421	_Ic13_6	1.02	0.97819
_Ipro_15	1.44	0.69628	b4	1.02	0.97910
Mean VIF	**1.99**				

Source: Authors' calculation based on 2002 CRPMS.

Household Head Characteristics. Male-headed households and age of the household head are negatively correlated with per capita consumption. This shows that male-headed households and head's age are contributory factors to increasing the number of poor. Interestingly, married household heads are more likely to be out of poverty than those who are not married.

Table 3.4 **Variance Inflation Factor of the OLS Regression Using the Data2 Subsample**

Variable	VIF	1/VIF	Variable	VIF	1/VIF
_Ib5_6	7.80	0.12818	c21	1.38	0.72622
_Ib5_3	6.98	0.14320	_Ipro_34	1.37	0.72877
_Ib5_4	6.81	0.14674	b22	1.35	0.74336
ln_p	5.31	0.18848	b19	1.33	0.75057
age0_14	4.05	0.24663	_Ipro_63	1.30	0.76988
age15_60	4.01	0.24911	b28	1.29	0.77374
_Ib5_2	3.96	0.25282	b47pc	1.28	0.77881
_Ipro_65	3.95	0.25332	a20	1.28	0.78034
_Ic13_3	3.79	0.26367	b26	1.26	0.79170
c7	3.51	0.28500	a6	1.26	0.79494
_Ic13_2	3.28	0.30470	_Ipro_64	1.25	0.80105
_Ipro_53	2.61	0.38265	fuel	1.25	0.80177
age60	2.40	0.41722	b23	1.23	0.81284
_Ib5_7	2.33	0.42994	b21	1.21	0.82877
laborr	2.29	0.43671	b31	1.17	0.85164
_Ic13_4	2.26	0.44185	b29	1.17	0.85285
studt	2.26	0.44340	_Ic13_5	1.17	0.85290
_Ib5_5	2.08	0.48185	c4	1.17	0.85681
ro_n_b10	1.99	0.50294	b72	1.16	0.86201
_Ipro_52	1.97	0.50793	b3	1.16	0.86441
landpc	1.83	0.54774	b17	1.16	0.86489
spouse	1.71	0.58535	a50	1.15	0.87159
_Ipro_45	1.70	0.58956	a57	1.14	0.87478
b20	1.65	0.60720	leadbus	1.14	0.87893
c5	1.61	0.61958	b18	1.13	0.88687
ro_n_b73	1.59	0.62696	_Ipro_46	1.13	0.88722
_Ipro_42	1.57	0.63705	b39	1.09	0.91404
b14	1.56	0.64043	b8	1.09	0.91454
_Ipro_41	1.56	0.64122	b34	1.09	0.91867
_Ipro_43	1.49	0.66998	cashr	1.07	0.93064
_Ipro_23	1.49	0.67229	b45pc	1.04	0.96378
_Ipro_15	1.46	0.68309	bigevent	1.04	0.96439
_Ipro_36	1.46	0.68456	b4	1.03	0.97133
_Ipro_50	1.45	0.68756	_Ic13_6	1.03	0.97352
_Ipro_14	1.45	0.69171	b46pc	1.02	0.98023
b13	1.40	0.71204	b25	1.02	0.98161
Mean VIF	**1.96**				

Source: Authors' calculation based on 2002 CRPMS.

In terms of education, a household with members with tertiary education or higher would have higher per capita expenditure and therefore is less likely to be poor compared with households whose members' level of education is low or nonexistent. This shows that gains from education in rural PRC can be manifested in the ability of the household head to provide for a higher standard of living.

Housing and Other Assets. Holding other factors constant, a household that has a telephone, truck, or TV usually has higher per capita expenditure and is less likely to be poor compared with a household that does not have these assets. Having a truck that can be used for economic activities, such as agricultural production, and having telephones and TVs suggests that a household can afford to spend on items beyond their basic needs.

However, having big animals (livestock) or sheep or goats could indicate for a lower per capita expenditure and the household with these assets is more likely to be poor compared with a household that does not have them. Typically, raising animals would imply savings due to the long gestation period of the animals. On the other hand, animals used for economic activities like a draught animal would increase the per capita consumption of the household.

In addition, a household that resides in larger houses and can store more grain has higher per capita consumption and is less likely to be poor. Other assets that suggest relatively nonpoor characteristics in a household are toilets, barns for livestock, and acreage.

Natural Resources. Land resources are positively correlated with household consumption, while environmental deterioration indicated by the difficulty of collecting fuels has a negative relationship with household consumption. Households engaged in large-scale agricultural production or business, or having family members who are village leaders or working outside the village, have a higher consumption level. In addition, households devoting more land to cash crops also have higher consumption.

Activities and Access to Services. Households that participate in insurance programs, use gas or coal for cooking, and have a big event taking place within the year also have higher consumption expenditures. However, households without any income sources (*Wu Bao Hu* in Chinese), participating in cooperative medical service, or having more family members staying at home have a lower consumption level.

A household that actively participates in community activities, such as being the village head or engaging in business, tends to consume more per household member and is less likely to be poor. High per capita consumption is also evident in big events such as weddings or funerals, or if the household has insurance. Expectedly, if the ratio of sown areas of cash crops to total sown areas in the community is higher, the household is less likely to be poor.

Community Characteristics. A number of community indicators are significantly correlated with household consumption. For instance, households living in villages designated as poor villages or those which encountered natural disasters have, as expected, low per capita consumption. Meanwhile, access to roads has also strong correlation with higher per capita consumption.

Predictability of the Ordinary Least Squares Method

To test the predicting capability of the ordinary least squares (OLS) models, Data1 was divided into three groups: bottom one-third, middle one-third and top one-third of the array of observations ranked according to actual and predicted per capita consumption expenditure. Table 3.5 shows that only 62 percent of the households that actually belong to the bottom one-third category were correctly predicted by the model, while the rest that were supposed to belong to the middle and top one-third were predicted to be under the bottom one-third category as well. Meanwhile, 43 percent of households in the middle one-third and 66 percent in the top one-third were correctly predicted by the model. Similar results can be observed when using Data2.

Table 3.5 **Accuracy of Predicted Expenditure**
Percent

Data1				
		Predicted		
		Bottom 33%	*Middle 33%*	*Top 33%*
Actual	*Bottom 33%*	62.15	30.11	7.73
	Middle 33%	30.11	43.27	26.63
	Top 33%	7.75	26.62	65.63

Data2				
		Predicted		
		Bottom 33%	*Middle 33%*	*Top 33%*
Actual	*Bottom 33%*	63.10	29.71	7.19
	Middle 33%	29.19	45.01	25.79
	Top 33%	7.70	25.28	67.03

Source: Authors' calculation based on 2002 CRPMS.

Likewise, to further test the predicting capability of the OLS model, households were divided into two groups, poor and nonpoor, depending on whether their per capita consumption expenditure was below or above the official poverty lines. With the low-income poverty line, about 51 percent of the households were predicted to be poor by the model, while almost 88 percent of the households were predicted to be nonpoor. Using the absolute poverty line, 98 percent of households were predicted to be nonpoor. The accuracy of predicting the poor was low at just 14 percent, indicating that it is very difficult to correctly predict the extreme poor using OLS regression (Tables 3.6 and 3.7). Again, similar results can be observed using Data2.

Logistic Regression Models

Summary results of the stepwise procedure for the logit model using the low-income poverty line for Data1 and Data2 were obtained (Table 3.8). As previously discussed, the Hosmer-Lemeshow test was used to test the goodness of fit of the model because some variables have sparse observations. The test revealed that the probability values are 0.4728 for Data1 and 0.1272 for Data2. Both statistics are lower than the expected probability, indicating that the models fit well with the data. See details of the results in Appendix 3.4–3.5.

The retained or significant variables in the logit regression after the stepwise procedure are almost the same with those of OLS regression but with opposite signs. This means that variables with negative coefficients would likely reduce the probability that a household is poor, and vice versa. Only a few variables that are significant in OLS regression are not significant in logit regression.

Table 3.6 **Accuracy of Predicted Poverty Status by Using the Low-Income Poverty Line**

Data1			
		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	87.55	12.45
	Poor	49.03	50.97

Data2			
		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	87.98	12.02
	Poor	49.15	50.85

Source: Authors' calculation based on 2002 CRPMS.

Table 3.7 **Accuracy of Predicted Poverty Status by Using the Absolute Poverty Line**

Data1			
		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	98.51	1.49
	Poor	85.79	14.21

Data2			
		Predicted	
		Nonpoor	*Poor*
Actual	*Nonpoor*	98.31	1.69
	Poor	85.29	14.71

Source: Authors' calculation based on 2002 CRPMS.

Table 3.8 **Summary Results of Stepwise Logit Regression for Model Building**

	Data1	**Data2**	**Absolute Poverty in Data1**
Number of observations	22,845	23,315	23,315
Hosmer-Lemeshow	7.61	12.58	8.06
Adjusted R-squared	0.4728	0.1272	0.4275

Note: Data1 and Data2 are subsamples of data set used for model building.
Source: Authors' calculation based on 2002 CRPMS.

Predictability of the Logit Method

To measure the accuracy of the prediction model, a number of indicators generated from the model were examined. Accuracy indicators vary with the choice of probability cutoff points. Table 3.9 shows the result taking 0.50

as the probability cutoff point while Table 3.9 shows the result taking 0.38 as the best probability cutoff point. The best cutoff point is determined by examining the sensitivity and specificity graph (Figure 3.4).

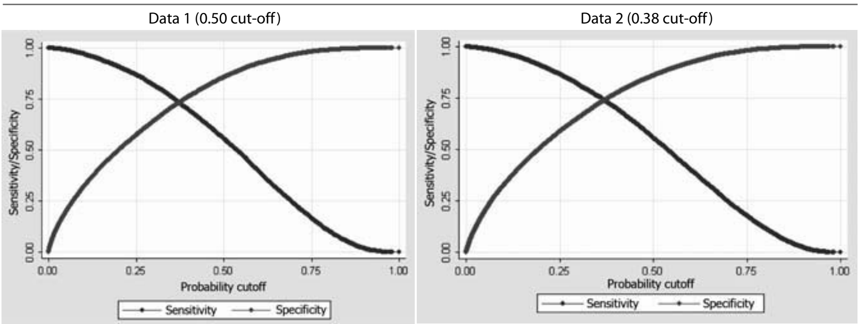
Table 3.9 **Accuracy of Predicted Poverty Status by Using Logit Regression and Low-Income Poverty Line**

	Probability Cutoff of 0.5 (Percent)		Probability Cutoff of 0.38 (Percent)	
	Data1	*Data2*	*Data1*	*Data2*
Sensitivity	55.59	55.73	72.09	72.61
Specificity	85.73	85.97	74.10	75.23
Positive predictive value	66.86	67.13	59.05	60.12
Negative predictive value	78.84	79.07	83.67	84.23
False positive rate for true nonpoor	14.27	14.03	25.90	24.77
False negative rate for true poor	44.41	44.27	27.91	27.39
False positive rate for classified poor	33.14	32.87	40.95	39.88
False negative rate for classified nonpoor	21.16	20.93	16.33	15.77
Correctly classified	75.44	75.70	73.41	74.34

Source: Authors' calculation based on 2002 CRPMS.

Table 3.9 shows that by using a probability cutoff of 0.50 and the low-income poverty line in Data1, about 56 percent percent of the poor households are correctly predicted (sensitivity), while 86 percent of nonpoor households are accurately predicted by the model (specificity). Positive predictive value measures the percentage of correctly predicted poor households to the total predicted poor households, while the negative predictive value measures the ratio of correctly predicted nonpoor to the total predicted nonpoor. The false positive rate for the true nonpoor indicates that 14 percent of nonpoor households are inaccurately predicted as poor households, while the false negative rate for the true poor indicates that 44 percent of poor households are inaccurately predicted as nonpoor households. The false positive rate for classified poor shows that 33 percent of the total predicted poor households are inaccurate, while 21 percent of the total predicted nonpoor households are not correct as shown by the false negative rate for classified nonpoor. The

Figure 3.4 **Sensitivity and Specificity of the Logit Regression**



Source: Authors' calculation.

overall accuracy of prediction is 75 percent. The general result for Data2 is again close to Data1.

Using the probability cutoff point of 0.38, on the other hand, reveals that the accuracy of poor household prediction is higher, that is, 72 percent, while the accuracy of nonpoor household prediction is less, that is, 74 percent. Meanwhile, the false prediction of the poor is less and the false prediction of the nonpoor is higher. The overall accuracy of prediction is also a little bit lower, that is 73 percent.

The stepwise procedure for the logit model is also implemented using the official absolute poverty line for Data1.[3] Table 3.10 reveals that, using the official absolute poverty line for defining the poverty status, only 17 percent of the poor households are correctly predicted if the 0.50 probability cutoff point was used. A simulation was also done using a different probability cutoff (Table 3.10). The simulation showed that prediction accuracy can increase by using a much lower probability cutoff point (0.16 in the simulation), but the false rate for predicting poor also increases (to a high of almost 70 percent in the simulation). The best cutoff point is determined by again examining the sensitivity and specificity graph in Figure 3.5. (See Appendix 3.6 for details.)

Table 3.10 **Accuracy of Predicted Poverty Status by Using Logit Regression and Official Absolute Poverty Line and Data 1**

	Probability Cutoff of 0.5	Probability Cutoff of 0.16
Sensitivity	17.41	73.17
Specificity	98.19	74.24
Positive predictive value	61.20	31.78
Negative predictive value	87.87	94.40
False positive rate for true non-poor	1.81	25.76
False negative rate for true poor	82.59	26.83
False positive rate for classified poor	38.80	68.22
False negative rate for classified non-poor	12.13	5.60
Correctly classified	86.80	74.09

Source: Authors' calculation based on 2002 CRPMS.

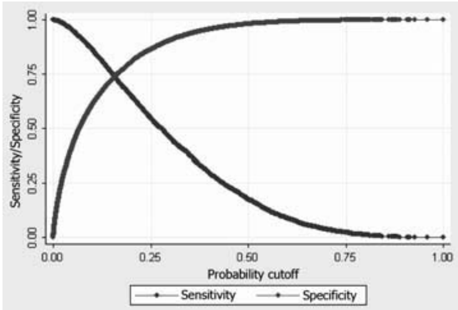
Summary and Conclusion

In the final selection of the poverty predictors, all independent variables that are significant in both OLS and logistic models were chosen. (See Appendix 3.7.)

Both the multiple linear regression models and the logistic regression model can accurately predict, by over 50 percent, which households are

3 The process was not conducted only for Data1 since the results of using Data 2 were negligibly different, as shown in previous results (See details in Appendix 3.8.).

Figure 3.5 **Sensitivity and Specificity of the Logit Regression Using the Absolute Poverty Line for Data1**



Source: Authors' calculation.

poor. The logistic regression model performs a little bit better than the OLS regression model in terms of predicting the poverty status of the households. Moreover, the logistic model is more flexible for choosing a probability cutoff point for higher prediction accuracy of the poor. The cost of doing so, however, is an increase of false prediction, which will lead to a spillover problem in program targeting. The modeling results show that predicting the extremely poor is very difficult.

To determine the accuracy of logit models for predicting which households are poor, the appropriate cutoff point is 0.38.

Appendix

Appendix 3.1 Candidate Variables Selected

Variable Name	Description
Welfare Indicators	
consumpc	Consumption expenditure per capita (yuan/person)
con_poor	Is the household consumption expenditure below the poverty line? 1=yes, 0=no
inc_poor	Is the household net income below the poverty line? 1=yes, 0=no
Household Head Characteristics	
C4	Sex of the household head, 1=male, 0=female
C5	Age of the household head
spouse	Whether the household head got married? 1=yes, 0=no
C7	Can household head speak Chinese? 1=yes, 0=no
C13	Education attainment of the household head
Household Demographics	
Age0_14	Number of family members aged 0–14 years
Age15_60	Number of family members aged 15–60 years
Age60	Number of family members over 60 years old
studt	Number of school age children in school
drops	Number of school age children dropped out of school
C16	Are there any disabled adults at home? 1=yes, 0=no
laborr	Ratio of labor to household members
B5	Family structure
Housing and Other Assets	
B13	Whether has big animals? 1=yes, 0=no
B14	Whether has pigs? 1=yes, 0=no
B15	Whether has sheep or goats? 1=yes, 0=no
B16	Whether has poultry? 1=yes, 0=no
B17	Whether has a radio? 1=yes, 0=no
B18	Whether has a refrigerator? 1=yes, 0=no
B19	Whether has a TV? 1=yes, 0=no
B20	Whether has a bicycle? 1=yes, 0=no
B21	Whether has a motorcycle? 1=yes, 0=no
B22	Whether has a telephone? 1=yes, 0=no
B25	Whether has a car or truck? 1=yes, 0=no
B26	Whether has a hand tractor? 1=yes, 0=no
B27	Whether has a large-or medium-sized tractor? 1=yes, 0=no
B28	Whether has a cart? 1=yes, 0=no
B29	Whether has other agricultural tools? 1=yes, 0=no
B30	Whether has a draught animal? 1=yes, 0=no
B31	Whether has a production animal? 1=yes, 0=no
B34	Whether has a toilet? 1=yes, 0=no
B72	Is grain enough for consumption? 1=yes, 0=no
n_b73	Grain stored at home at the end of the year (kg/person)
n_b75	Grain stored for consumption at home at the end of the year (kg/person)
NB12	Whether the house is built with bricks or concrete? 1=yes, 0=no
n_b10	Square meters of living house per capita
B23	Square meters of production (business) house
B24	Square meters of barn for livestock
Natural Resources	
landpc	Cultivated land per capita, mu/per person
B45pc	Forest land per capita (mu/person)
B46pc	Orchard land per capita (mu/person)
B47pc	Grassland areas per capita (mu/person)
B48pc	Water areas under cultivation per capita (mu/person)
B49pc	Wasteland areas per capita (mu/person)
B39	Whether is it difficult to access drinking water? 1=yes, 0=no
B41	Whether it become more difficult to collect fuels? 1=yes, 0=no
Activities and Access to Services	
n_p	Number of household members staying at home for 6 months or more
B3	Whether engaged in large-scale agricultural production? 1=yes, 0=no
leadbus	Is any family members the village leader or engaged in business? 1=yes, 0=no
C21	Are there any household members who work outside? 1=yes, 0=no
cashr	Ratio of sown areas of cash crop to total sown areas
fuel	Whether use coal or gas for cooking? 1=yes, 0=no
B4	Whether a “wu bao hu” without any income sources, 1=yes, 0=no
B6	Whether participated in cooperatives? 1=yes, 0=no
B7	Whether participated in cooperative medical service? 1=yes, 0=no
B8	Whether has insurance? 1=yes, 0=no
C6	Does the household belong to ethnic minority groups? 1=yes, 0=no
B35	Whether has electricity? 1=yes, 0=no
bigevent	Whether has a big event such as wedding, funeral, etc. 1=yes, 0=no
Community Characteristics	
A1	Village physiognomy
A6	Number of natural villages with a road for motor vehicles
A14	Distance to the countryseat, km
A15	Distance to the town where the township government locates, km
A20	Distance to the nearby market, km
A50	Whether had a natural disaster in the village? 1=yes, 0=no
A57	Whether being designated as a poor village? 1=yes, 0=no

Source: Based on Household Survey Questionnaire.

Appendix 3.2 **Results of Stepwise Ordinary Least Square Regression Using Data1**
(Dependent Variable: Log Per Capita Expenditure)

Variable Name	Description	Coefficient	Standard Error	P>|t|
Household Demographics				
age0_14	Number of family members aged 0–14 years old	0.047	0.006	0.000
age15_60	Number of family members aged 15–60 years old	0.104	0.005	0.000
age60	Number of family members over 60 years old	0.095	0.007	0.000
studt	Number of school age children in school	0.077	0.004	0.000
_Ib5_2	Households with a couple and one child	0.175	0.016	0.000
_Ib5_3	Households with a couple and two children	0.229	0.017	0.000
_Ib5_4	Households with a couple and three children or more	0.216	0.019	0.000
_Ib5_5	Households with father or mother and the children	0.206	0.025	0.000
_Ib5_6	Households with three generations	0.242	0.019	0.000
_Ib5_7	Other kinds of households	0.210	0.023	0.000
Household Head Characteristics				
c4	Sex of the household head	-0.066	0.017	0.000
c5	Age of the household head	-0.001	0.000	0.001
spouse	Whether the household head got married?	0.122	0.015	0.000
c7	Can household head speak Chinese?	0.089	0.019	0.000
_Ic13_2	Household head with primary school education	0.041	0.011	0.000
_Ic13_3	Household head with middle school education	0.084	0.012	0.000
_Ic13_4	Household head with high school education	0.112	0.014	0.000
_Ic13_5	Household head with technical secondary school education	0.181	0.029	0.000
_Ic13_6	Household head with college education and above	0.309	0.088	0.000
Housing and Other Assets				
ro_n_b10	Square root of housing acreage	0.037	0.003	0.000
b23	Square meters of production (business) house	0.000	0.000	0.007
b24	Square meters of barn for livestock	0.001	0.000	0.001
b13	Whether has big animals?	-0.045	0.011	0.000
b15	Whether has sheep or goats?	-0.034	0.009	0.000
b17	Whether has a radio?	0.020	0.007	0.004
b18	Whether has a refrigerator?	0.075	0.015	0.000
b19	Whether has a TV?	0.094	0.008	0.000
b20	Whether has a bicycle?	0.022	0.007	0.004
b21	Whether has a motorcycle?	0.086	0.010	0.000
b22	Whether has a telephone?	0.146	0.009	0.000
b25	Whether has a truck?	0.093	0.032	0.004
b26	Whether has a hand tractor?	0.035	0.009	0.000
b30	Whether has a draught animal?	0.038	0.011	0.001
b31	Whether has a production animal?	0.036	0.008	0.000
b34	Whether has a toilet?	0.062	0.025	0.013
ro_n_b73	Square root of the amount of grain stored at home per capita	0.004	0.000	0.000
Natural Resources				
b41	Whether it becomes more difficult to collect fuels?	-0.030	0.007	0.000
landpc	Cultivated land per capita	0.007	0.001	0.000
b45pc	Forest land per capita	0.007	0.001	0.000
b47pc	Grassland areas per capita	0.000	0.000	0.000
Activities and Access to Services				
ln_p	Log of family members staying at home for 6 months or more	-0.936	0.017	0.000
b3	Whether engaged in large-scale agricultural production?	0.057	0.018	0.002
leadbus	Is any family member the village leader or engaged in business?	0.089	0.011	0.000
c21	Any household members working outside?	0.088	0.008	0.000
cashr	Ratio of sown areas of cash crop to total sown areas	0.139	0.017	0.000
fuel	Whether use coal or gas for cooking?	0.032	0.007	0.000
b4	Whether a "wu bao hu" without any income sources	-0.150	0.061	0.014
b7	Whether participated in cooperative medical service?	-0.040	0.019	0.041
b8	Whether has insurance?	0.060	0.010	0.000
bigevent	Whether has a big event?	0.195	0.008	0.000
Community Characteristics				
_Ia1_2	Hilly areas	0.022	0.008	0.006
a6	Number of natural villages with a road for motor vehicles	0.002	0.001	0.022
a15	Distance to the town where the township government is located	0.001	0.000	0.033
a20	Distance to the nearby market	0.002	0.000	0.000
a50	Whether had a natural disaster in the village?	-0.034	0.007	0.000
a57	Whether designated as a poor village?	-0.047	0.006	0.000
Provincial Dummy				
_Ipro_14	Shanxi	-0.086	0.014	0.000
_Ipro_15	Inner Mongolia	0.103	0.017	0.000
_Ipro_22	Jilin	-0.060	0.026	0.022
_Ipro_34	Anhui	0.177	0.017	0.000
_Ipro_36	Jiangxi	0.240	0.017	0.000
_Ipro_41	Henan	0.112	0.014	0.000
_Ipro_42	Hubei	0.288	0.016	0.000
_Ipro_43	Hunan	0.299	0.017	0.000
_Ipro_45	Guangxi	0.308	0.016	0.000

(continued on next page)

Appendix 3.2 continued

Variable Name	Description	Coefficient	Standard Error	P>|t|
_Ipro_46	Hainan	0.284	0.037	0.000
_Ipro_50	Chongqing	0.271	0.019	0.000
_Ipro_52	Guizhou	0.223	0.014	0.000
_Ipro_53	Yunnan	0.155	0.013	0.000
_Ipro_63	Qinghai	0.340	0.025	0.000
_Ipro_64	Ningxia	0.144	0.026	0.000
_Ipro_65	Xinjiang	0.291	0.023	0.000
_cons		6.974	0.053	0.000

Number of obs = 22845
F(72, 22772) = 273.58
Prob > F = 0.0000
Adj R-squared = 0.4621

P |t| = probability of accepting the null hypothesis (Ho)
Source: Authors' calculation based on 2002 CRPMS.

Appendix 3.3 **Results of Stepwise Ordinary Least Square Regression Using Data2**
(Dependent Variable: Log Per Capita Expenditure)

Variable Name	Description	Coefficient	Standard Error	P>|t|
Household Demographics				
age0_14	Number of family members aged 0–14 years old	0.032	0.006	0.000
age15_60	Number of family members aged 15–60 years old	0.096	0.005	0.000
age60	Number of family members over 60 years old	0.068	0.007	0.000
Studt	Number of school age children in school	0.076	0.004	0.000
_Ib5_2	Households with a couple and one child	0.154	0.016	0.000
_Ib5_3	Households with a couple and two children	0.197	0.017	0.000
_Ib5_4	Households with a couple and three children or more	0.186	0.019	0.000
_Ib5_5	Households with father or mother and the children	0.143	0.025	0.000
_Ib5_6	Households with three generations	0.221	0.019	0.000
_Ib5_7	Other kinds of households	0.187	0.023	0.000
laborr	Ratio of labor to household members	-0.064	0.019	0.001
Household Head Characteristics				
c4	Sex of the household head	-0.045	0.017	0.008
c5	Age of the household head	-0.001	0.000	0.011
spouse	Whether the household head got married?	0.106	0.015	0.000
c7	Can household head speak Chinese?	0.075	0.021	0.000
_Ic13_2	Household head with primary school education	0.039	0.011	0.000
_Ic13_3	Household head with middle school education	0.086	0.011	0.000
_Ic13_4	Household head with high school education	0.114	0.014	0.000
_Ic13_5	Household head with technical secondary school education	0.216	0.028	0.000
_Ic13_6	Household head with college education and above	0.239	0.071	0.001
Housing and Other Assets				
ro_n_b10	Square root of housing acreage	0.030	0.003	0.000
b23	Square meters of production (business) house	0.001	0.000	0.000
b13	Whether has big animals?	-0.014	0.007	0.044
b14	Whether have pigs?	0.032	0.008	0.000
b17	Whether has a radio?	0.034	0.007	0.000
b18	Whether has a refrigerator?	0.039	0.014	0.006
b19	Whether has a TV?	0.103	0.008	0.000
b20	Whether has a bicycle?	0.037	0.007	0.000
b21	Whether has a motorcycle?	0.095	0.009	0.000
b22	Whether has a telephone?	0.123	0.008	0.000
b25	Whether has a truck?	0.133	0.032	0.000
b26	Whether has a walking tractor?	0.020	0.009	0.036
b28	Whether has a cart?	-0.027	0.010	0.007
b29	Whether have other agricultural tools?	0.049	0.008	0.000
b31	Whether has a production animal?	0.033	0.008	0.000
b34	Whether has a toilet?	0.082	0.022	0.000
ro_n_b73	Square root of amount of grain stored at home per capita	0.004	0.000	0.000
Natural Resources				
b39	Whether is it difficult to access drinking water?	-0.018	0.008	0.019
landpc	Cultivated land per capita	0.009	0.001	0.000
b45pc	Forest land per capita	0.001	0.001	0.039
b46pc	Orchard land per capita	0.020	0.006	0.001
b47pc	Grassland areas per capita	0.001	0.000	0.000
Activities and Access to Services				
ln_p	Log of family members staying at home for 6 months or more	-0.933	0.017	0.000
b3	Whether engaged in large-scale agricultural production?	0.104	0.018	0.000
leadbus	Is any family members the village leaders or engaged in business?	0.087	0.010	0.000
c21	Any household members working outside?	0.091	0.007	0.000
cashr	Ratio of sown areas of cash crop to total sown areas	0.104	0.017	0.000
b72	Is self-produced grain enough for consumption?	0.035	0.009	0.000
fuel	Whether use coal or gas for cooking?	0.041	0.007	0.000
b4	Whether a "wu bao hu" without any income sources	-0.175	0.060	0.003
b8	Whether has insurance?	0.061	0.010	0.000
bigevent	Whether has a big event?	0.186	0.008	0.000
Community Characteristics				
a6	Number of natural villages with road for motor vehicles	0.002	0.001	0.001
a20	Distance to the nearby market	0.002	0.000	0.000
a50	Whether had a natural disaster in the village?	-0.035	0.006	0.000
a57	Whether designated as a poor village?	-0.018	0.006	0.003
Provincial Dummy				
_Ipro_14	Shanxi	-0.034	0.015	0.021
_Ipro_15	Inner Mongolia	0.101	0.017	0.000
_Ipro_23	Heilongjiang	0.053	0.021	0.011
_Ipro_34	Anhui	0.223	0.017	0.000
_Ipro_36	Jiangxi	0.303	0.017	0.000
_Ipro_41	Henan	0.147	0.014	0.000
_Ipro_42	Hubei	0.388	0.016	0.000
_Ipro_43	Hunan	0.352	0.017	0.000
_Ipro_45	Guangxi	0.320	0.016	0.000
_Ipro_46	Hainan	0.289	0.037	0.000

(continued on next page)

Appendix 3.3 continued

Variable Name	Description	Coefficient	Standard Error	P>|t|
_Ipro_50	Chongqing	0.278	0.019	0.000
_Ipro_52	Guizhou	0.237	0.014	0.000
_Ipro_53	Yunnan	0.175	0.013	0.000
_Ipro_63	Qinghai	0.311	0.025	0.000
_Ipro_64	Ningxia	0.088	0.026	0.001
_Ipro_65	Xinjiang	0.338	0.024	0.000
_cons		6.873	0.038	0.000

Number of obs = 23115
F(72, 23042) = 282.63
Prob > F = 0.0000
Adj R-squared = 0.4673

Source: Authors' calculation based on 2002 CRPMS.

Appendix 3.4 **Results of Stepwise Logit Regression Using Data1**
(Dependent Variable: Poor = 1, Nonpoor= 0)

Variable Name	Description	Coefficient	Standard Error	P>|z|
Household Demographics				
age0_14	Number of family members aged 0–14 years old	-0.173	0.038	0.000
age15_60	Number of family members aged 15–60 years old	-0.377	0.032	0.000
age60	Number of family members over 60 years old	-0.346	0.044	0.000
studt	Number of school age children in school	-0.320	0.023	0.000
_Ib5_2	Households with a couple and one child	-0.762	0.096	0.000
_Ib5_3	Households with a couple and two children	-1.052	0.101	0.000
_Ib5_4	Households with a couple and three childern or more	-1.008	0.114	0.000
_Ib5_5	Households with father or mother and the children	-0.859	0.149	0.000
_Ib5_6	Households with three generations	-1.178	0.115	0.000
_Ib5_7	Other kinds of households	-1.028	0.130	0.000
Household Head Characteristics				
c5	Age of the household head	0.007	0.002	0.000
spouse	Whether the household head got married?	-0.363	0.080	0.000
c7	Can household head speak Chinese?	-0.535	0.112	0.000
_Ic13_3	Household head with middle school education	-0.179	0.038	0.000
_Ic13_4	Household head with high school education	-0.338	0.063	0.000
_Ic13_5	Household head with technical secondary school education	-0.332	0.166	0.045
_Ic13_6	Household head with college education and above	-1.601	0.763	0.036
Housing and Other Assets				
ro_n_b10	Square root of housing acreage	-0.154	0.017	0.000
b23	Square meters of production (business) house	-0.004	0.001	0.000
b15	Whether has sheep or goats?	0.220	0.050	0.000
b17	Whether has a radio?	-0.109	0.038	0.005
b18	Whether has a refrigerator?	-0.214	0.090	0.018
b19	Whether has a TV?	-0.384	0.043	0.000
b21	Whether has a motorcycle?	-0.391	0.058	0.000
b22	Whether has a telephone?	-0.555	0.052	0.000
b26	Whether has a hand tractor?	-0.107	0.052	0.040
b31	Whether has a production animal?	-0.182	0.042	0.000
b35	Whether has electricity?	-0.169	0.084	0.043
ro_n_b73	Square root of the amount of grain stored at home per capita	-0.028	0.004	0.000
ro_n_b75	Square root of the amount of grain stored at home for consumption per capita	0.009	0.004	0.047
Natural Resources				
b39	Whether is it difficult to access drinking water?	0.122	0.043	0.005
b41	Whether it becomes more difficult to collect fuels?	0.107	0.037	0.004
landpc	Cultivated land per capita	-0.040	0.007	0.000
b45pc	Forest land per capita	-0.046	0.012	0.000
b47pc	Grassland areas per capita	-0.009	0.001	0.000
b49pc	Wasteland areas per capita	-0.091	0.022	0.000
Activities and Access to Services				
ln_p	Log of family members staying at home for 6 months or more	3.803	0.142	0.000
leadbus	Is any family members the village leaders or engaged in business?	-0.398	0.066	0.000
c21	Any household members working outside?	-0.509	0.044	0.000
Cashr	Ratio of sown areas of cash crop to total sown areas	-0.616	0.099	0.000
b72	Is self-produced grain enough for consumption?	0.107	0.049	0.030
Fuel	Whether use coal or gas for cooking?	-0.226	0.041	0.000
b7	Whether participated in cooperative medical service?	0.239	0.103	0.020
b8	Whether has insurance?	-0.239	0.060	0.000
bigevent	Whether has a big event?	-0.515	0.045	0.000
Community Characteristics				
a6	Number of natural villages with a road for motor vehicles	-0.011	0.004	0.008
a15	Distance to the town where the township government is located	-0.007	0.002	0.002
a50	Whether had a natural disaster in the village?	0.196	0.037	0.000
a57	Whether designated as a poor village?	0.199	0.035	0.000
Provincial Dummy				
_Ipro_14	Shanxi	0.348	0.077	0.000
_Ipro_15	Inner Mongolia	-0.395	0.098	0.000
_Ipro_23	Heilongjiang	-0.303	0.116	0.009
_Ipro_34	Anhui	-0.730	0.100	0.000
_Ipro_36	Jiangxi	-1.493	0.113	0.000
_Ipro_41	Henan	-0.460	0.077	0.000
_Ipro_42	Hubei	-1.351	0.102	0.000
_Ipro_43	Hunan	-1.362	0.099	0.000
_Ipro_45	Guangxi	-1.288	0.090	0.000
_Ipro_46	Hainan	-1.344	0.194	0.000
_Ipro_50	Chongqing	-1.277	0.116	0.000
_Ipro_52	Guizhou	-0.984	0.073	0.000
_Ipro_53	Yunnan	-0.558	0.066	0.000
_Ipro_63	Qinghai	-1.199	0.142	0.000
_Ipro_64	Ningxia	-0.468	0.143	0.001
_Ipro_65	Xinjiang	-1.415	0.134	0.000
_cons		-0.316	0.209	0.130

number of observations = 22845
number of groups = 10
Hosmer-Lemeshow chi2(8) = 7.61
Prob > chi2 = 0.4728

Source: Authors' calculation based on 2002 CRPMS.

Appendix 3.5 Results of Stepwise Logit Regression Using Data2
(Dependent Variable: Poor = 1; Nonpoor = 0)

Variable Name	Description	Coefficient	Standard Error	P>|z|
Household Demographics				
age0_14	Number of family members aged 0–14 years old	-0.090	0.038	0.018
age15_60	Number of family members aged 15–60 years old	-0.309	0.032	0.000
age60	Number of family members over 60 years old	-0.171	0.048	0.000
Studt	Number of school age children in school	-0.338	0.023	0.000
c16	Are there any disabled adults at home?	-0.118	0.051	0.020
_Ib5_2	Households with a couple and one child	-0.687	0.095	0.000
_Ib5_3	Households with a couple and two children	-0.909	0.099	0.000
_Ib5_4	Households with a couple and three children or more	-0.850	0.113	0.000
_Ib5_5	Households with father or mother and the children	-0.619	0.144	0.000
_Ib5_6	Households with three generations	-1.012	0.113	0.000
_Ib5_7	Other kinds of households	-0.831	0.131	0.000
Household Head Characteristics				
c4	Sex of the household head	0.198	0.099	0.046
c5	Age of the household head	0.004	0.002	0.037
Spouse	Whether the household head got married?	-0.354	0.083	0.000
_Ic13_2	Household head with primary school education	-0.197	0.058	0.001
_Ic13_3	Household head with middle school education	-0.422	0.062	0.000
_Ic13_4	Household head with high school education	-0.535	0.079	0.000
_Ic13_5	Household head with technical secondary school education	-0.829	0.183	0.000
Housing and Other Assets				
ro_n_b10	Square root of housing acreage	-0.118	0.017	0.000
b23	Square meters of production (business) house	-0.004	0.001	0.000
b13	Whether has big animals?	0.078	0.039	0.047
b14	Whether have pigs?	-0.203	0.044	0.000
b17	Whether has a radio?	-0.152	0.038	0.000
b19	Whether has a TV?	-0.471	0.042	0.000
b20	Whether has a bicycle?	-0.191	0.043	0.000
b21	Whether has a motorcycle?	-0.352	0.057	0.000
b22	Whether has a telephone?	-0.553	0.051	0.000
b25	Whether has a truck?	-0.461	0.194	0.018
b26	Whether has a hand tractor?	-0.122	0.053	0.022
b28	Whether has a cart?	0.129	0.057	0.022
b29	Whether have other agricultural tools?	-0.265	0.050	0.000
b31	Whether has a production animal?	-0.157	0.043	0.000
b34	Whether has a toilet?	-0.427	0.151	0.005
ro_n_b73	Square root of the amount of grain stored at home per capita	-0.021	0.003	0.000
Natural Resources				
landpc	Cultivated land per capita	-0.045	0.007	0.000
b45pc	Forest land per capita	-0.035	0.014	0.014
b46pc	Orchard land per capita	-0.292	0.075	0.000
b47pc	Grassland areas per capita	-0.005	0.001	0.000
Activities and Access to Services				
ln_p	Log of family members staying at home for 6 months or more	3.572	0.141	0.000
b3	Whether engaged in large-scale agricultural production?	-0.303	0.105	0.004
leadbus	Is any family member the village leader or engaged in business?	-0.385	0.065	0.000
c21	Any household members working outside?	-0.581	0.044	0.000
cashr	Ratio of sown areas of cash crop to total sown areas	-0.323	0.100	0.001
b72	Is self-produced grain enough for consumption?	-0.124	0.049	0.011
fuel	Whether use coal or gas for cooking?	-0.197	0.041	0.000
b4	Whether a “wu bao hu” without any income sources	0.658	0.323	0.042
b8	Whether has insurance?	-0.235	0.058	0.000
bigevent	Whether has a big event?	-0.540	0.046	0.000
Community Characteristics				
_Ia1_3	Mountainous areas	-0.098	0.044	0.025
a20	Distance to the nearby market	-0.007	0.002	0.000
a50	Whether had a natural disaster in the village?	0.190	0.036	0.000
a57	Whether designated as a poor village?	0.076	0.035	0.028
Provincial Dummy				
_Ipro_14	Shanxi	0.296	0.077	0.000
_Ipro_15	Inner Mongolia	-0.495	0.099	0.000
_Ipro_23	Heilongjiang	-0.425	0.116	0.000
_Ipro_34	Anhui	-1.022	0.106	0.000
_Ipro_36	Jiangxi	-1.574	0.112	0.000
_Ipro_41	Henan	-0.528	0.081	0.000
_Ipro_42	Hubei	-1.704	0.107	0.000
_Ipro_43	Hunan	-1.747	0.103	0.000
_Ipro_45	Guangxi	-1.148	0.090	0.000
_Ipro_46	Hainan	-1.358	0.197	0.000
_Ipro_50	Chongqing	-1.279	0.116	0.000
_Ipro_52	Guizhou	-1.001	0.079	0.000
_Ipro_53	Yunnan	-0.696	0.068	0.000
_Ipro_63	Qinghai	-0.992	0.140	0.000
_Ipro_65	Xinjiang	-1.130	0.093	0.000
_cons		0.131	0.218	0.548

Number of observations = 23115
Hosmer-Lemeshow chi2(8) = 12.58
Prob > chi2 = 0.1272

Source: Authors’ calculation based on 2002 CRPMS.

Appendix 3.6 **Results of Stepwise Logit Regression Using the Absolute Poverty Line and Dataset1** (Dependent Variable: Poor = 1, Nonpoor = 0)

Variable Name	Description	Coefficient	Standard Error	P>|z|
Household Demographics				
age15_60	Number of family members aged 15–60 years old	-0.238	0.027	0.000
age60	Number of family members over 60 years old	-0.180	0.052	0.001
Studt	Number of school age children in school	-0.314	0.028	0.000
Drops	Number of school age children dropped out of school	0.179	0.075	0.018
c16	Are there any disabled adults at home? 1=yes, 0=no	-0.129	0.065	0.046
_Ib5_2	Households with a couple and one child	-0.689	0.136	0.000
_Ib5_3	Households with a couple and two children	-0.927	0.101	0.000
_Ib5_4	Households with a couple and three children or more	-0.898	0.152	0.000
_Ib5_5	Households with father or mother and the children	-0.790	0.120	0.000
_Ib5_6	Households with three generations	-0.999	0.154	0.000
_Ib5_7	Other kinds of households	-0.770	0.172	0.000
Household Head Characteristics				
c5	Age of the household head	0.007	0.002	0.002
Spouse	Whether the household head got married?	-0.255	0.099	0.010
c7	Can household head speak Chinese?	-0.347	0.127	0.006
_Ic13_3	Household head with middle school education	-0.268	0.050	0.000
_Ic13_4	Household head with high school education	-0.290	0.087	0.001
Housing and Other Assets				
ro_n_b10	Square root of housing acreage	-0.162	0.023	0.000
b24	Square meters of barn for livestock	-0.008	0.001	0.00
b14	Whether have pigs?	-0.125	0.056	0.026
b15	Whether has sheep or goats?	0.136	0.062	0.029
b19	Whether has a TV?	-0.468	0.053	0.000
b21	Whether has a motorcycle?	-0.362	0.080	0.000
b22	Whether has a telephone?	-0.671	0.076	0.000
b26	Whether has a hand tractor?	-0.198	0.070	0.005
b27	Whether has a large or medium sized tractor? 1=yes, 0=no	0.333	0.137	0.015
B28	Whether has a cart? 1=yes, 0=no	0.146	0.068	0.031
b35	Whether has electricity?	-0.344	0.095	0.000
ro_n_b73	Square root of the amount of grain stored at home per capita	-0.030	0.004	0.000
Natural Resources				
b39	Whether is it difficult to access drinking water?	0.161	0.054	0.003
b41	Whether it becomes more difficult to collect fuels?	0.130	0.048	0.007
Landpc	Cultivated land per capita	-0.072	0.010	0.000
b45pc	Forest land per capita	-0.066	0.021	0.002
b47pc	Grassland areas per capita	-0.014	0.003	0.000
b49pc	Wasteland areas per capita	-0.160	0.043	0.000
Activities and Access to Services				
ln_p	Log of family members staying at home for 6 months or more	3.128	0.144	0.000
leadbus	Is any family members the village leaders or engaged in business?	-0.283	0.092	0.002
c21	Any household members working outside?	-0.606	0.059	0.000
Cashr	Ratio of sown areas of cash crop to total sown areas	-0.505	0.129	0.000
b4	Whether a "wu bao hu" without any income sources, 1=yes, 0=no	0.942	0.363	0.010
bigevent	Whether has a big event?	-0.389	0.060	0.000
Community Characteristics				
a20	Distance to the nearby market, km	-0.009	0.002	0.000
a50	Whether had a natural disaster in the village?	0.245	0.049	0.000
a57	Whether designated as a poor village?	0.232	0.045	0.000
Provincial Dummy				
_Ipro_14	Shanxi	0.205	0.092	0.026
_Ipro_15	Inner Mongolia	-0.568	0.145	0.000
_Ipro_34	Anhui	-1.191	0.161	0.000
_Ipro_36	Jiangxi	-1.904	0.198	0.000
_Ipro_41	Henan	-0.440	0.105	0.000
_Ipro_42	Hubei	-1.586	0.167	0.000
_Ipro_43	Hunan	-2.046	0.172	0.000
_Ipro_45	Guangxi	-1.763	0.141	0.000
_Ipro_46	Hainan	-1.739	0.292	0.000
_Ipro_50	Chongqing	-1.785	0.207	0.000
_Ipro_52	Guizhou	-1.497	0.111	0.000
_Ipro_53	Yunnan	-0.699	0.095	0.001
_Ipro_62	Gansu	-0.304	0.094	0.000
_Ipro_63	Qinghai	-1.359	0.192	0.000
_Ipro_64	Ningxia	-0.879	0.197	0.000
_Ipro_65	Xinjiang	-1.629	0.167	0.000
_cons		-0.727	0.296	0.014

number of observations = 22819

number of groups = 10

Hosmer-Lemeshow chi2(8) = 8.06

Prob > chi2 = 0.4275

Source: Authors' calculation based on 2002 CRPMS.

Appendix 3.7 Identified Poverty Predictors

Variable Name	Description
Household Demographics	
age0_14	Number of family members aged 0–14 years old
age15_60	Number of family members aged 15–60 years old
age60	Number of family members over 60 years old
studt	Number of school age children in school
c16	Are there any disabled adults at home? 1=yes, 0=no
laborr	Ratio of labor to household members
b5	Family structure
Household Head Characteristics	
c4	Sex of the household head, 1=male, 0=female
c5	Age of the household head
spouse	Whether the household head got married? 1=yes, 0=no
c7	Can household head speak Chinese? 1=yes, 0=no
c13	Education attainment of the household head
Housing and Other Assets	
n_b10	Square meters of housing per capita
b23	Square meters of production (business) house
b24	Square meters of barn for livestock
b13	Whether has big animals? 1=yes, 0=no
b14	Whether has pigs? 1=yes, 0=no
b15	Whether has sheep or goat? 1=yes, 0=no
b17	Whether has a radio? 1=yes, 0=no
b18	Whether has a refrigerator? 1=yes, 0=no
b19	Whether has a TV? 1=yes, 0=no
b20	Whether has a bicycle? 1=yes, 0=no
b21	Whether has a motorcycle? 1=yes, 0=no
b22	Whether has a telephone? 1=yes, 0=no
b25	Whether has a car or truck? 1=yes, 0=no
b26	Whether has a hand tractor? 1=yes, 0=no
b28	Whether has a cart? 1=yes, 0=no
b29	Whether has other agricultural tools? 1=yes, 0=no
b30	Whether has a draught animal? 1=yes, 0=no
b31	Whether has a production animal? 1=yes, 0=no
b34	Whether has a toilet? 1=yes, 0=no
b35	Whether has electricity? 1=yes, 0=no
b72	Is grain enough for consumption? 1=yes, 0=no
n_b73	Grain stored at home at the end of the year (kg/person)
n_b75	Grain stored for consumption at home at the end of the year (kg/person)
Natural Resources	
landpc	Cultivated land per capita, mu/per person
b45pc	Forest land per capita (mu/person)
b46pc	Orchard land per capita (mu/person)
b47pc	Grassland areas per capita (mu/person)
b49pc	Wasteland areas per capita (mu/person)
b39	Whether is it difficult to access drinking water? 1=yes, 0=no
b41	Whether it becomes more difficult to collect fuels? 1=yes, 0=no
fuel	Whether use coal or gas for cooking? 1=yes, 0=no
Activities and Access to Services	
b3	Whether engaged in large scale agricultural production? 1=yes, 0=no
Leadbus	Is any family members the village leaders or engaged in business? 1=yes, 0=no
n_p	Number of household members staying at home for 6 months or more
c21	Are there any household members who work outside? 1=yes, 0=no
Cashr	Ratio of sown areas of cash crop to total sown areas
b4	Whether a “wu bao hu” without any income sources, 1=yes, 0=no
b7	Whether participated in cooperative medical service? 1=yes, 0=no
b8	Whether has insurance? 1=yes, 0=no
bigevent	Whether has a big event such as wedding, funeral, etc. 1=yes, 0=no
Community Characteristics	
a1	Village physiognomy
a6	Number of natural villages with a road for motor vehicles
a15	Distance to the town where the township government is located, km
a20	Distance to the nearby market, km
a50	Whether had a natural disaster in the village? 1=yes, 0=no
a57	Whether being designated as a poor village? 1=yes, 0=no
pro	Provincial code

Source: Authors’ calculation based on 2002 CRPMS.

CHAPTER 4

Poverty Predictor Modeling in the People's Republic of China: A Validation Survey

Pingping Wang

Introduction

Based on poverty predictors identified in Sangui, Pingping, and Heng (2005) and listed in Appendix 3.1, a short questionnaire was developed and used in a pilot survey to determine whether or not the poor in a particular location could be identified without conducting an income and expenditure survey. If the tool could be used to identify the poor, it would be useful for evaluating the impact of a poverty reduction project on a target area. To be able to validate the results of the survey, the questionnaire included questions on the respondents' income and expenditures. A comparison was also carried out on the accuracy of the assessment of households' poverty status based on results of different assessors.

Data and Methods

Sample Size and Data Gathering

The pilot survey[1] was conducted in five counties in the province of Yunnan in the People's Republic of China (PRC). The coverage area was along the Asian Development Bank–financed Kunming-Dali expressway. A total of 1,000 households spread over 50 villages were interviewed. In each county, there were 10 villages and 200 households selected. In each village, 20 households were selected, of which 10 households were from the sample coverage of the China Rural Poverty Monitoring Survey (CRPMS), while the rest were newly selected samples. A total of 45 villages with 450 households were taken from the CRPMS while 5 villages and 550 households were non-CRPMS.

Field supervisors had made several trips to check and ensure that the enumerators followed the guidelines of the survey manual, directly assess the

[1] The questionnaire used in the pilot survey can be downloaded at http://www.adb.org/Statistics/reta_6073.asp.

poverty status of the households according to the poverty predictors, observe the reaction of respondents to the survey questions, and discuss the survey with government staff of counties and townships, village heads, villagers, owners and employees of enterprises, farmers, etc.

The pilot survey also identified the poverty status of households based on judgments of village heads, neighbors, enumerators, and the households themselves.

Income and living expenditure data were collected through daily recording and were regarded as *actual* data in this study. The result was compared with the perception of household poverty status based on the independent assessments.

Validation Method

As a preliminary step, the significance of the predictors of household poverty status was first validated using the results from the pilot survey data and the existing national poverty monitoring survey, that is, the CRPMS. The coefficients of poverty predictors of the ordinary least squares (OLS) model for the subsample group Data1 in Sangui, Pingping, and Heng (2005) were applied to 450 sample households from the CRPMS to predict the per capita living expenditure for the said sample. The result was regarded as *predicted* data in this study.

Next, the levels of predicted and actual per capita expenditure were compared with poverty lines CNY700,[2] CNY1,000, and CNY1,500 to determine the measures of poverty status. CNY700 was an approximation of the official rural poverty line, which was CNY668 in 2004. CNY1,000 was an approximation of the current official poverty line for the low-income group, which was CNY924 in 2004 and was about $1-a-day at purchasing power parity prices. Finally, CNY1,500 was an approximation of the proposed poverty line for the rural upper-income group. Also, data were divided into low-, middle-, and high-income groups based on per capita expenditure and predicted and actual data were compared. Cross tabulation of actual and predicted poverty measures as well as income groups would reveal the accuracy of the poverty predictors.

The next task was to build the new OLS regression and logit models using the results of the pilot survey and the significant predictor variables previously mentioned. For OLS regression, predicted per capita consumption derived from the survey was then compared to the three poverty lines mentioned above to again determine the measures of poverty status. Actual and predicted measures were again cross tabulated to reveal accuracy. For the logit model,

2 CNY stands for Chinese Yuan.

sensitivity and specificity coefficients were directly computed to determine the accuracy of the prediction.

In eliminating the bias of self-reporting, the respondent's welfare status was also evaluated by three other individuals: village head, the respondent's neighbor, and the survey enumerator. The respondent was rated by evaluators according to the following categories: poor, low-income, and nonpoor.

For the final step of validation, means of measures of poverty predictors for poor and nonpoor were subjected to a test of mean difference using a t-test.

Results

Poverty-Predictor Accuracy Based on 450 CRPMS Households

Applying the coefficients of poverty predictors of the OLS model to 450 sample households from the CRPMS would reveal that expected value of per capita consumption is quite close to the actual daily reporting of individual consumption with minimum variance (Table 4.1).

Table 4.1 **Statistical Summaries of Per Capita Expenditure**

Variable	*Number*	*Mean (CNY)*	*Standard Error*
Actual	450	1664.57	1180.49
Predicted	450	1673.26	615.26

Source: Authors' calculation based from 2002 CRPMS.

As shown in Table 4.2, as the poverty line increases, the accuracy of predicting the poor household increases, while the reverse is observed in predicting the nonpoor. It might be noted that everyone with per capita consumption above CNY700, is predicted as nonpoor, which implies that there could be serious prediction problems if the poverty line used is too low. This is in line with the finding of this book's Chapter 3.

Table 4.2 **Poverty Status Using the CNY700, CNY1000, and CNY1500 Poverty Lines—Actual Versus Predicted**

		Predicted				
		700 CNY	*1000 CNY*		*1500 CNY*	
		Nonpoor	*Nonpoor*	*Poor*	*Nonpoor*	*Poor*
Actual	*Nonpoor*	100.0	98.5	1.5	73.2	26.8
	Poor	100.0	88.1	11.9	44.7	55.3

Source: Authors' calculation based on 2002 CRPMS.

To further validate the model, the households' per capita expenditure was divided into low, middle, and high groups.[3] The empirical result shows that poverty among the low-income group can be predict ed at 61 percent, while the high-income group can only be predicted at 59 percent. The middle group seems to have low prediction capability (Table 4.3).

Table 4.3 **Comparing Households Based on Per Capita Expenditure—Actual Versus Predicted**

		Predicted			
		Low	*Middle*	*High*	*Total*
Actual	*Low*	61.30	28.70	10.00	100.00
	Middle	22.70	46.00	31.30	100.00
	High	16.00	25.30	58.70	100.00
	Total	100.00	100.00	100.00	-

Source: Authors' calculation based on 2002 CRPMS.

Poverty Predictor Accuracy of Households in the Pilot Survey

From the OLS estimation, the model generated predicted per capita expenditures, which were then compared with the three poverty lines. As shown in Table 4.4, increasing poverty lines increase the likelihood of accurately predicting the poor but the reverse is observed in predicting the nonpoor.

Table 4.4 **Classifying Poor and Nonpoor Using the Per Capita Expenditure—Actual Versus Predicted**

		Predicted Based on Per Capita Living Expenditure					
		700 CNY		*1000 CNY*		*1500 CNY*	
		Nonpoor	*Poor*	*Nonpoor*	*Poor*	*Nonpoor*	*Poor*
Actual	*Nonpoor*	98.8	1.20	91.0	9.0	72.1	27.9
	Poor	68.8	31.30	59.0	41.0	23.5	76.5

Source: Authors' calculation based on 2002 CRPMS.

Logistic regression was also used to predict whether a household was poor or not. Here, poverty was measured using CNY1,500 per capita expenditure as the poverty line. The dependent variable was whether the household was poor (with per capital expenditure below CNY1,500), where *1* is poor and *0* is nonpoor.

Accordingly, as shown in Table 4.5, the percentage of poor correctly predicted was about 82 percent and the percentage of nonpoor correctly predicted was around 76 percent. This indicates that logistic regression is more powerful than OLS regression in terms of predicting poverty. The

3 All households were divided equally based on predicted per capita consumption as well as actual per capita consumption.

probability of incorrectly predicting the poor (poor that were actually not poor), is 24 percent while the probability of the opposite case is 18 percent.

Table 4.5 **Accuracy of Predicted Poverty Status Using the Logit Model with CNY1,500 Poverty Line**
(percent)

Sensitivity	82.04
Specificity	76.14
Positive predictive value	80.09
Negative predictive value	78.36
False positive rate for true nonpoor	23.86
False negative rate for true poor	17.96
False positive rate for classified poor	19.91
False negative rate for classified nonpoor	21.64
Correctly classified	79.32

Probability cut off of 0.20
Source: Authors' calculation based on 2002 CRPMS.

An Alternative Approach for Identifying the Poor

Using the evaluators' judgment of the respondents' poverty status, results reveal that while the respondents themselves perceive that most of them belong to low-income or poor groups, the evaluators perceive the respondents to be in low-income or nonpoor groups (Table 4.6). Thus, there was an upward bias in estimating the number of poor based on respondents' own perceptions.

Table 4.6 **Classification of Poor and Nonpoor Based on Different Assessors**
(percent)

Assessors	*Poor*	*Low-Income*	*Nonpoor*	*Total*
Village head	7.50	20.60	71.90	100.00
Enumerator	5.50	19.40	75.10	100.00
Neighbor	7.50	20.70	71.80	100.00
Respondent: based on income	10.70	76.70	12.60	100.00
Respondent: based on expenditure	19.40	74.20	6.40	100.00

Source: Authors' calculation based on 2002 CRPMS.

Using the 1,000 household responses, the local perception of poverty was matched with the identified poverty predictors. A respondent was categorized as poor if and only if all evaluators rated the respondent as such. If the respondent rated himself or herself as poor and the rest of the evaluators did not, the respondent was classified as nonpoor. This method classified 138 households as poor category, while 119 households were classified as nonpoor. The predictors were considered to be reliable if they were present in poor households but not in nonpoor households.

Table 4.7 shows the mean values of the poverty predictor variables from the survey results. The last column shows the t-Statistics of the differences in the means of the nonpoor and poor. A predictor was eliminated if the difference was not significantly different from 0 at a 95 percent confidence level, that is, when both poor and nonpoor households were locally perceived to have the same characteristics.

For further refinement, those that did not provide substantial information on the differences between poor and nonpoor were also eliminated. For instance, the average number of residents per household for the nonpoor was 4.56 and for the poor it was 4.22. Although their t-statistic for mean difference was high enough, the predictor does not notably distinguish between the two groups.

Table 4.7 also shows that some identified poverty predictors that have positive coefficients from the linear regression model developed in Sangui, Pingping, and Heng (2005)–indicating that the higher value of the predictor increases the log of per capita expenditure of a household–turned out to be more apparent among poor households than in nonpoor ones. Family structure, where the household has other members apart from immediate family, is an example of such a poverty indicator. The coefficient for the linear regression was positive when only 5 percent among the nonpoor households have other members, whereas it was 14 percent among the poor households.

The new sets of predictors provide indicators of the household's poverty status. Of the 1,000 households, 15 percent have at least one of the demographic characteristics, 84 percent possess at least one of the assets common to poor households, 99 percent have heads that were either single or have a high school education or less (up to none at all), and 21 percent live in mountainous areas. There were only 42 households that met all of the four criteria above and almost half of them were identified to be poor by at least one of the evaluators.

Table 4.8 presents the percentage distribution of households classified as poor according to the group of predictors. Notable is the high percentage (83 percent) of the population that were categorized as poor because they have at least one of the assets common to poor households and have household heads that are either single or have low education levels. There was a small percentage of the population who were classified as poor because of their household demographics and because they live in mountainous areas.

Table 4.7 **Mean of Poverty Predictors and T-Statistics of the Mean Difference**

Household Characteristics	PPM Coefficient +/-	Mean		t-Statistics
		Nonpoor	Poor	
Household Demographics				
Number of residents		4.56	4.22	2.10
Aged 0–14 years	+	1.49	1.40	0.94
Aged 15–60 years	+	3.31	2.86	3.21
Aged over 60 years old	+	1.26	1.32	-0.57
Staying at home for 6 months or more	-	4.19	4.12	0.39
Number of school-age children in school	+	1.48	1.42	0.59
Family structure:				
Has parents and no children	+	0.03	0.00	1.45
Has parents and one child	+	0.13	0.13	0.09
Has parents and two children	++	0.27	0.29	-0.34
Has parents and three children or more	++	0.03	0.00	1.45
Has either one of the parents and children	++	0.00	0.06	-2.50
Has three generations	++	0.45	0.34	1.72
Has other members	++	0.05	0.14	-2.32
Has disabled adults at home	ns	0.02	0.19	-4.62
Ratio of labor to household members	-	0.67	0.61	2.32
Activities and Access to Services				
Celebrates big events	++	0.21	0.27	-1.05
Engaged in large-scale production	+	0.05	0.02	1.21
A household member is the village leader	+	0.28	0.03	5.60
Number of members that work outside the village	+	1.53	1.26	1.88
Ratio of cash crop areas to total sown areas	+	0.26	0.23	0.92
Has grain that is enough for consumption	+	0.99	0.94	2.28
Uses coal or gas for cooking	+	0.65	0.28	6.25
Has no income sources (Wu Bao Hu)	-	0.00	0.00	-
Participates in cooperative medical service	-	0.06	0.00	2.48
Has insurance	+	0.37	0.11	5.00
Asset Ownership				
Has big animals	-	0.69	0.65	0.65
Has pigs	+	0.68	0.90	-4.53
Has sheep or goat	-	0.04	0.18	-3.68
Has a radio	+	0.44	0.25	3.25
Has a refrigerator	+	0.19	0.02	4.46
Has a TV	+	0.99	0.67	7.76
Has a bicycle	+	0.72	0.29	7.49
Has a motorcycle	+	0.28	0.07	4.52
Has a telephone	+	0.63	0.18	8.12
Has a car or truck	+	0.11	0.00	3.61
Has a hand tractor	+	0.06	0.02	1.40
Has other agricultural tools	+	0.26	0.29	-0.65
Has draught animal	+	0.38	0.59	-3.38
Has production animal	+	0.40	0.24	2.69
Has toilet	+	0.91	0.68	4.96
Has electricity	ns	1.00	0.97	2.02
Amount of grain stored at home at the end of the year (kg/person)	+	332.40	295.24	1.45

(continued on next page)

Table 4.7 continued

Household Characteristics	PPM Coefficient +/-	Mean Nonpoor	Mean Poor	t-Statistics
Amount of grain stored for consumption at home at the end of the year (kg/person)	ns	220.18	165.02	3.05
Floor area of house per household member (square meters)	+	36.37	31.52	2.12
Area of house allotted for production (square meters)	+	51.37	46.60	0.76
Area of barn for livestock (square meters)	ns	34.06	29.10	1.76
Has difficult access to drinking water	-	0.11	0.34	-4.44
Finds collecting fuels getting more difficult	-	0.47	0.61	-2.34
Natural Resources				
Area of cultivated land per capita	+	1.16	1.05	1.50
Area of forest land per capita	+	1.61	2.36	-0.91
Area of orchard land per capita	ns	0.40	0.40	-0.02
Area of grassland areas per capita	+	0.15	0.10	1.29
Wasteland areas per capita	ns	1.06	0.77	0.42
Household Head Characteristics				
Sex of the household head is male		0.92	0.93	-0.32
Age of the household head	-	44.77	42.57	1.70
Marital status:				
Single	-	0.01	0.10	-2.98
Married	+	0.96	0.83	3.70
Divorce		0.01	0.06	-2.00
Household head can speak Chinese	+	0.99	0.99	-0.10
Educational attainment:				
Without formal education	+	0.01	0.12	-3.49
With primary school education	+	0.33	0.54	-3.40
With middle school education	+	0.52	0.29	3.85
With high school education	+	0.10	0.20	2.30
With college education or higher	++	0.01	0.00	0.68
Village Characteristics				
Village physiognomy:				
Has plate land	+	0.60	0.47	2.04
Has hilly areas	+	0.32	0.06	5.45
Has mountainous areas	ns	0.06	0.45	-8.04
Number of natural villages with a road for motor vehicles	+	10.47	15.97	-5.43
Distance to the town where the township government is located (km)	+	2.13	2.74	-4.52
Distance to the nearby market (kilometers)	+	2.44	2.80	-2.59
Natural disaster occurs in the village	-	0.85	0.52	5.85
Village designated as poor by the National Poverty Reduction Project	-	0.37	0.15	4.01

ns = not (statistically) significant
Source: Authors' calculation based on the household survey used by Sangui, Pingping, and Heng.

Table 4.8 **Distribution of Households Identified as Poor**
(Percent)

Identified Poor by:	*Household Demographics*	*Asset Ownership*	*Household Head Characteristics*	*Village Characteristics*
Household Demographics	14.7	11.7	14.7	4.4
Asset Ownership	11.7	83.5	83.0	20.5
Household Head Characteristics	14.7	83.0	99.3	20.9
Village Characteristics	4.4	20.5	20.9	21.1

Source: Authors' calculation based on the household survey with N=1,000 households as generated by Sangui, Pingping, and Heng.

Conclusion

Although every country's poverty situation is unique, the underlying determinants of poverty generally point to a household having low income or facing limited access to income sources. The poverty predictors generated in this study suggest that households are poor because they either have low income or difficult access to income sources. The first can be attributed to having fewer income earners, which was evident form the poor households' characteristics. The second can be attributed to the households' inability to generate higher income because of low education levels that limit them from engaging in other gainful economic activities, or the households' geographic location that prevents them from having access to wider markets for their products and services.

In addition, some predictors, such as those under asset ownership, were outcomes rather than determinants of income status. For instance, a household with a radio, refrigerator, TV, bicycle, motorcycle, telephone, among other assets, was generally classified as nonpoor. Poor households, on the other hand, generally have sheep or goats, or have difficulty accessing drinking water and fuel. The capability of households to purchase relatively more expensive assets signify higher income compared with those who cannot afford them. On the other hand, the inability of households to acquire easier access to drinking water, for instance, signifies lower income compared with those who can afford household appliances.

The poverty predictors thus covered indicators of both causes and effects of poverty. Because the predictors were initially derived by correlating the household's per capita consumption expenditure and the household's characteristics, they reflect the relevance of purchasing power as a factor in defining poverty. In addition, because they were also derived using local perceptions of poverty, the predictors likewise reflect the multidimensional aspects of poverty that include not only the level of income but also other factors that make a household socially and economically disadvantaged.

The households classified as poor by community characteristics, for instance, were poor because they were located in mountainous areas and were not able to generate as much farm income as those households located on flatter land. The cost of living in mountainous regions is usually higher and, hence, some of the households classified as nonpoor by a common poverty line may in fact be poor in this region. The predictors, therefore, go beyond the numeric definition of poverty set by poverty lines.

In terms of the accuracy of the poverty predictor model, the empirical study suggests that the logistic regression model is more accurate than the multiple regression technique. With the given set of predictors or variables to characterize the poor and nonpoor, a survey is an effective instrument to monitor and evaluate the impact of poverty-related projects in the PRC. However, for the purpose of evaluating the effectiveness of the project, the identified poverty predictor variables should be incorporated in the instrument before the start of any poverty reduction project or program.

CHAPTER 5

Identifying Poverty Predictors Using Household Living Standards Surveys in Viet Nam

Linh Nguyen

Introduction

Poverty predictor modeling (PPM) based on a regression-type analysis of household income and expenditure and other variables (predictors) from household surveys of living standards, has been receiving more attention from researchers and practitioners. This interest comes from the fact that PPM provides an easy and low-cost way to collect baseline and follow-up poverty measures for monitoring progress and evaluating the poverty impact of development projects and policies. But while PPM is popular, the reliability of this methodology has yet to be checked.

In Viet Nam, there have been a number of efforts to develop and use poverty predictor models for poverty mapping (Minot 1998, Minot and Baulch 2002 and 2003, MOLISA 2005). These studies were mostly intended for use in poverty targeting and budget transfers. There has been no effort, however, to apply the approach to ex-ante poverty estimates of participatory assessments of various policies. Moreover, there has been no attempt to use data sets of the subsequent comparable household surveys to assess how good the predictors really are.

The approach presented in this study is an attempt to develop a practical alternative to the time-consuming and expensive collection of income and expenditure data for assessing poverty at local levels. In Phase 1 of the study, data from 2002 living standards surveys of Viet Nam's General Statistical Office were used to examine the relationship between poverty and a household's characteristics using a multiple regression modeling technique. This technique detects variables or predictors that have correlated effects on a household's living standards and, consequently, its poverty status. In Phase 2, significant predictors were tested using a 1997/98 living standards survey to check the consistency and stability of the models across time. In Phase 3, another regression modeling procedure was implemented for two provinces in the North Central Coast subregion to further test the methodology and to check whether the poverty predictors would be different at more a disaggregated level. Finally, in Phase 4, reliable and easy-to-collect poverty predictors within the regression model were used to generate a short questionnaire[1] for frequent implementation or for data collection at local levels.[2]

Data and Methods

Data

For Phases 1 and 2, the work uses the 1997/98 Viet Nam Living Standard Survey (VLSS) and the 2002 Viet Nam Household Living Standard Survey (VHLSS), both implemented by the General Statistical Office. These surveys provide data on income, expenditure, and other characteristics of households such as demography, education, health, assets, housing, etc. They are fairly well-organized, have high-quality data, and can be a good source of information for poverty analysis and assessment at the national and even at the provincial levels.

The 2002 VHLSS data were crucial to this work. The information was used to derive the basic poverty predictor model and to test the stability of the model. The survey had a general sample size of 75,000 households and collected information about household living standards and basic communal socioeconomic conditions including income and expenditures. Income data came from all 75,000 households, but expenditure data were from only 30,000 households.

The total sample used in the study was composed of 29,510 households. For comparison, the sample was split into urban and rural data sets. There were 22,601 rural households in the sample, while the rest were urban. To test the stability of the model across the whole data set, the rural and urban data sets were further split into a learning data set and a validation data set. This was done by randomly drawing a subsample of 50 percent of the total sample as the learning data set for both rural and urban areas. The other 50 percent subsample was used as the validation data set. The learning and validation data sets had to be very similar to each other to ensure the comparability of the two models' statistics. Summary statistics of the 2002 VHLSS rural data set are presented in Table 5.1.

1 The questionnaire used in the pilot survey can be downloaded at http://www.adb.org/Statistics/reta_6073.asp.

2 Aside from predictors, some questions were also included in the questionnaire to create variables for specific studies relating to poverty.

Method for Phase 1

Table 5.1 **Summary Statistics of the 2002 Viet Nam Household Living Standard Survey of Rural Area**

Variable	Samples	Mean	Standard Deviation
Learning	11,299	2,838.758	1,672.116
Validation	11,302	2,842.604	1,633.516

Source: Author's calculation.

The Model. The ultimate goal of this study was to build a good regression model to examine the relationship between household expenditure and household characteristics using the 2002 VHLSS. Multiple regression modeling was the method employed in the study in the following form:

$$\textit{Dependent Variable} = \beta_0 + (\textit{Independent Variable}_i \times \beta_i) + e_i$$

The dependent variable was the household's annual expenditure per capita or one of its transformations, rather than income as a measure of household living standards, to ensure international comparability.[3] The right-hand side variables were household characteristics from survey data, also called poverty predictors. The model's parameters were as follows: β_0 was the model intercept or constant, while β_i were respective regression coefficients. Finally, e_i were random errors that included effects of all variables on the dependent variable other than the ones explicitly considered in the model.

The commonly used method, weighted least squares, was used in this study to estimate model parameters (β_0 and β_i) by minimizing the sum of random errors e_i across households using the sampling weight. It worked by incorporating extra nonnegative constants or weights associated with each data point into the fitting criterion. The size of the weight indicated the precision of the information contained in the associated observation.

Optimizing the weighted fitting criterion to find the parameter estimates allowed the use of weights to determine the contribution of each observation to the final parameter estimates. It was important to note that the weight for each observation was given relative to the weights of the other observations; so different sets of absolute weights could have identical effects.[4]

A model-building procedure was implemented on the learning data set until a satisfactory model of poverty predictors was achieved. Next, the predictor variables were created based on the validation data set, which was in turn used as a basis for creating the poverty predictor model. Finally, the statistics of the two models for the learning and validation data sets were compared. If these statistics were similar, then the model was considered

3 Income is usually more underestimated than expenditure in household surveys, which is another reason for using expenditure in the model.

4 See http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd143.htm.

stable across the data set. If they were not similar, the whole process would be repeated for another regression model for the learning data set until the model statistics for the two data sets were similar.

Hence, model building was done for four subsamples: urban and rural areas, both disaggregated by learning and validation data sets. The model was first constructed for the rural subsample, then the same procedure was applied for the urban subsample.

Variable Selection. For the dependent variable, the choice was between annual expenditure per capita and some of its transformations. A number of transformations such as natural logarithm, logarithm, square root, etc., were generated and examined. The natural logarithm of annual per capita expenditure (log of PCE) was eventually selected as the dependent variable since this type of transformation most closely follows the normal distribution.

For independent variables, a list was created for all possible variables using household characteristics that were believed to affect household living standards. From the 2002 VHLSS household questionnaire, 60 variables of this type were chosen including region, household size, number of household members under or above certain ages, household assets (black-and-white TV, colored TV, rice cooker, motorbike, etc.), occupation of the head, and number of unemployed members. Many variables relating to households' agricultural activities such as number and proportion of people working in agriculture and size of land areas were also used since these activities were very important aspects in the lives of people in rural areas. Since the aim of the study was to predict the dependent variable and not to estimate the determinants (causality) of household living standards, the endogeneity of the independent variables was not a concern.

From the list of independent variables, only easy-to-collect variables were chosen to meet the requirement of creating a short questionnaire (which was built in Phase 2) that could be completed quickly. These independent variables were examined carefully to create an overview or metadata of mean, minimum, and maximum values, and to see if a variable was categorical or continuous, among other things (see Appendix 5.1 for the list of variables). Dummies were used during the model-building process which increased the number of variables to more than 60.

To examine and narrow down the number of variables, tests were conducted in three stages. First, a bivariate data analysis was done in which each independent variable was evaluated based on the strength of its individual relationship with the log of PCE. Variables with a significant relationship with the dependent variable were retained. The analysis used an F-test for means for categorical variables (see Table 5.2 for an example) and a correlation coefficient test for continuous variables (see Table 5.3 for an example).[5] Both tests selected variables that generated probability values less than the assigned significant level. Selected variables that were highly correlated with the dependent variable were retained in the model.

Table 5.2 **Example of F-Test for Means Using the Categorical Variables**

Obs	Categorical Variable	Sample Size	DF	SS1	F-stat	Prob
1	motorbike	11,297	1	264575.8	2421.92	0.0000000
2	colortv (color tv)	11,297	1	251205.9	2274.88	0.0000000
3	ricecooker (rice cooker)	11,297	1	245796.6	2216.29	0.0000000
4	gascooker (gas cooker)	11,297	1	243019.5	2186.40	0.0000000
5	telephone	11,297	1	197464.4	1714.35	0.0000000
6	toilet	11,292	6	298012.4	467.12	0.0000000
7	num_u15 (household member under 15 years old)	11,290	8	248647.7	280.71	0.0000000
8	num_dep (number of dependent)	11,289	9	227154.0	224.08	0.0000000
9	refee (rental fee)	11,297	1	176345.6	1506.55	0.0000000
...	...	...	...	...	...	...

Obs = observation; DF = Degrees of freedom; SS = Sum of squares; F-stat = Statistics; Prob = Probability of acceptance
Source: Authors' calculation based on 2002 VLSS.

Table 5.3 **Example of Correlation Coefficient Test for Continuous Variables**

Pearson Correlation Coefficients, N = 11299					
Prob > |r| under H0: Rho=0					
Dv	prop_u15	prop_o15	livingarea	prop_dep	prop_labor
Corr. Coef.	-0.35539	0.35539	0.23516	-0.20947	0.20947
Prob	<.0001	<.0001	<.0001	<.0001	<.0001
Dv	prop_illi	hage	prop_o60	prop_o70	prop_studmem
Corr. Coef.	-0.17242	0.13166	0.09637	0.05286	-0.00678
Prob	<.0001	<.0001	<.0001	<.0001	0.4713

Note: prop_u15 = Proportion of household members under 15 years; leavingarea = Leaving area; prop_dep = proportion of dependents; prop_labor = proportion of persons in the labor force (15–16 years); prop_illi = proportion of illiterate people; hage = age of household head; prop_o60 = proportion of member where age = 60; prop_o70 = proportion of member where age = 70; prop_studmem = proportion of studying people
Source: Authors' calculation based on 2002 VLSS.

The second stage in selecting variables involved a multivariate analysis on multicollinearity between predictors. Some of the independent variables

5 A continuous variable has numeric values such as 1, 2, 3, 4, 5, etc. The relative magnitude of the values is significant. For example, a value of 2 indicates twice the magnitude of 1. On the other hand, a categorical variable, also known as a nominal variable, has values that function as labels rather than as numbers. For example, a categorical variable for gender might use the value *1* for male and *2* for female; marital status might be coded as *1* for single, *2* for married, *3* for divorced, and *4* for widowed. Some software applications allow the use of nonnumeric (character-string) values for categorical variables. Hence, a data set could have the strings *Male* and *Female* or *M* and *F* for a categorical gender variable. Because categorical values are stored and compared as string values, a categorical value of *001* is different from the value of *1*. In contrast, values of *001* and *1* would be equal for continuous variables (see http://www.dtreg.com/vartype.htm).

could have been highly correlated with each other and, therefore, would have been redundant. This redundancy could have caused problems in the modeling process. In the multivariate analysis, a correlation test was run for pairs of independent variables. If the correlation coefficient of two independent variables was equivalent to 80 percent and above, then it was assumed that multicollinearity existed between these two variables. However, even if there was multicollinearity, variables that had a high degree of relationship with the dependent variables were kept (see Appendixes 5.2, 5.3, and 5.6 for the list of candidate variables).

The final stage in selecting the variables involved transforming continuous independent variables. For this purpose, the variables chosen from the previous stage were plotted against the log of PCE. In Figure 5.1, the shapes of the plot suggest independent variables should be transformed. Possible transformations were also tested in conjunction with the dependent variable (see Table 5.4 for an example). The transformed variables that generated high correlation were retained. Table 5.5 lists the variables that were transformed in this study.

Table 5.4 **Transformation of Nonlinear Independent Variables to Minimize Error**

Variables	Transformation
Urban file	
• proportion of dependent people (prop_dep)	Truncated at 90th percentile
• proportion of people studying (prop_studmen)	Square root
• proportion of people 15 years old or older (prop_o15)	Square root
Rural file	
• proportion of dependent people (prop_dep)	Square root
• proportion of illiterate people (prop_illi)	Square root
• age of household head (hage)	Natural logarithm
• agricultural land area (agriland)	Natural logarithm

Source: Author's summary based on the modeling development results.

A test for multicollinearity was again done to track down possible multicollinearity among transformed and untransformed variables. From this test, the list of the best candidate variables was finalized for use in the model-building process.

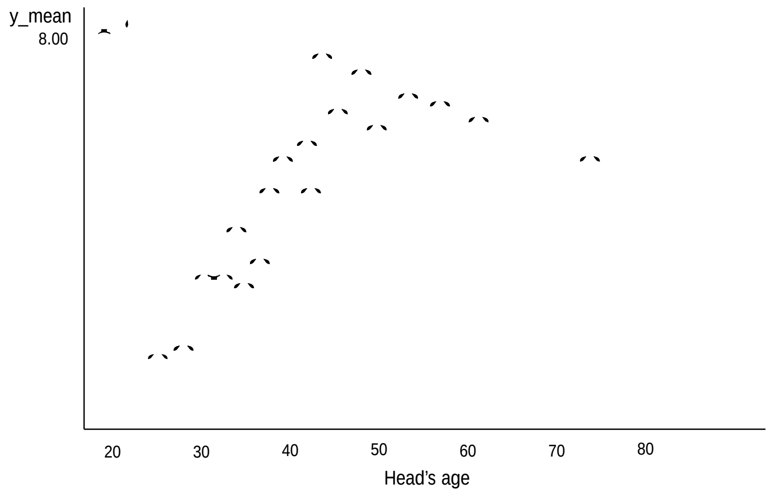
Table 5.5 **Transformation of Nonlinear Independent Variables**

Pearson Correlation Coefficients, N = 4822
Prob > |r| under H0: Rho=0

	Transformation Type				
	Natural Logarithm	*Square Root*	*Truncated at 95th percentile*	*Truncated at 99th percentile*	*No transformation*
Correlation coefficient	0.03712	0.03198	0.03031	0.02745	0.02643
Probability	0.0099	0.0264	0.0353	0.0567	0.0665

Independent Variable: Head's age
Source: Author's calculation based on 2002 VLSS.

Figure 5.1 **Example of Variable Plot that Needs Transformation**



Note: The scatter plot suggest a curvilinear or non-linear that has to be transformed to satisfy linearity criteria for the model.
Source: Author's calculation.

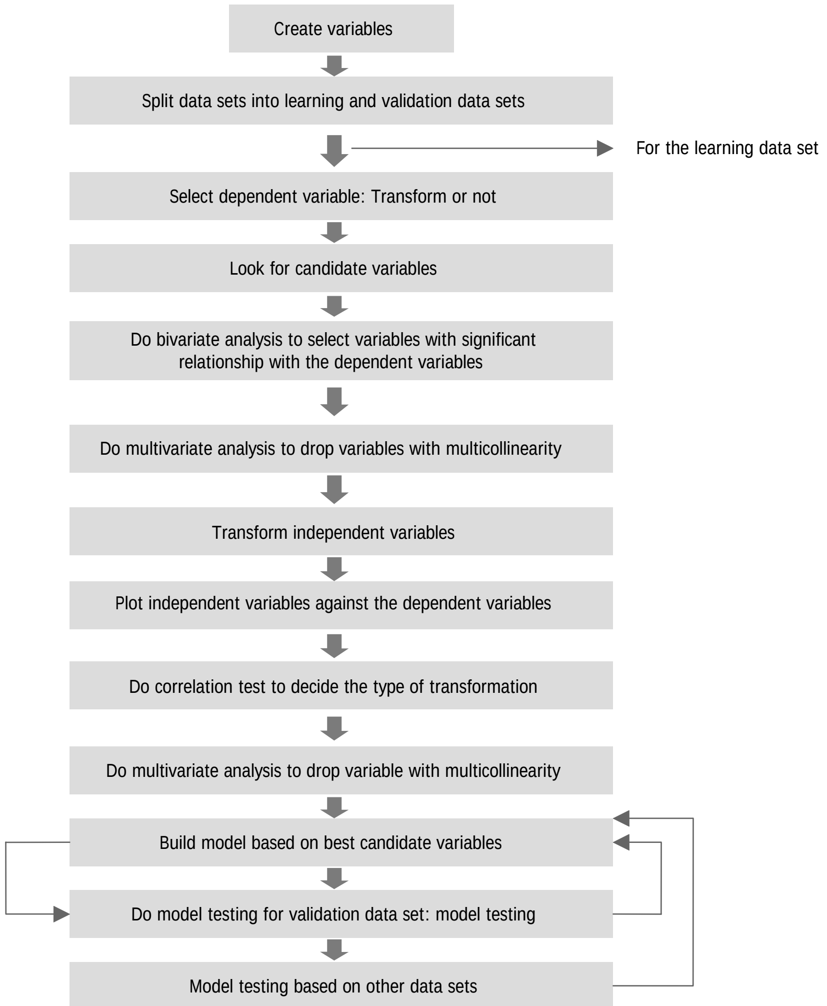
Model Building. The model was built using the learning data set for rural and urban areas, and weighted using the sample weight of the survey. Model-adequacy checks were performed by examining the R-squared values, residual plot, and plot of actual versus predicted values of log PCE for constancy of variance test and matched tabulation to see if top and bottom quintiles were balanced.

As mentioned in a previous section, subsamples for rural and urban areas were each split into learning and validation data sets to test the stability of the model across the subsamples. The model created using the learning data set would be applied to the validation data set. The following were the criteria considered for developing the model:

- The same set of predictors were significant in the validation model.
- The correlation direction of these predictors was the same as the dependent variable.
- Model statistics for the two data sets were similar or negligibly different.

Figure 5.2 is a summary of the steps in the methodology.

Figure 5.2 **Flow Chart for Building a Poverty Predictor Model**



Source: Author's framework.

Method for Phase 2

To further ensure that the final model was the best model possible, significant predictors were tested and validated using the 1997/98 VLSS.[6] The test was

6 The 1992/93 VLSS, the General Statistical Office's earliest living standards survey, was not considered in the study because data were too old to be used for testing the model.

to examine the stability of the model across time. All the model statistics and selection criteria were also reviewed for this model to see how much the chosen predictors fit in the 1997/98 VLSS. The 1997/98 VLSS collected information on 6,000 households. It does not include income data but, like the 2002 VHLSS, it gathered more detailed information on household expenditure, household characteristics, and commune data.

Method for Phase 3

To further test the methodology or disprove that poverty predictors may be different when estimating for a more disaggregated level than the national level, another regression modeling procedure was implemented for two provinces in the North Central Coast subregion, namely, Thanh Hoa and Nghe An, using the 2002 VHLSS. The selected subregion accounted for the biggest share of rural poor households in the country based on the 2002 VHLSS. While constructing the poverty predictor model for Thanh Hoa and Nghe An, two variables were added to the list of candidate variables, that is, *maize* (households harvesting maize = 1) and *sugarcane* (households harvesting sugarcane = 1) since these agricultural products are popular and indigenous crops in these provinces. Data sets were also equally split into learning and validation subsamples to test the stability of the whole data set, each with only 705 observations.

Method for Phase 4

After the identification of the variables necessary for the poverty predictor model, a pilot survey was implemented. The main objective was to assess the effectiveness of the poverty predictor model in estimating the poverty rate of the subregion taking into consideration the perceptions of respondents themselves (self-assessment), enumerators, and hamlet chiefs on household poverty classification. The survey used a questionnaire that contains not only variables identified in the poverty predictor model, but also questions on the interventions that the government or international organizations provided and could provide, as well as emerging issues on trade liberalization.

The sampling method used in this pilot survey was the two-stage cluster random sampling. The survey was conducted in Thanh Hoa and Nghe An with a sample size of 500 households. The results of the 2004 VHLSS were used as a benchmark in assessing the effectiveness of the survey, specifically, in classifying poor households. The results of the 2004 VHLSS were also used as a sampling frame for the pilot survey.

Results in Phases 1 and 2

Rural Areas

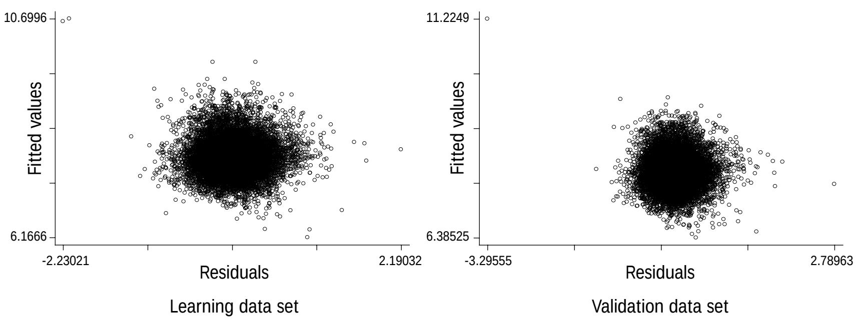
In general, the results for the rural areas were acceptable as shown in Table 5.6. The model from the learning data set generated an R-squared of 0.5801; for the validation data set, the R-squared was 0.5762. In other words, about 58 percent of the changes in the log of PCE was due to changes in the retained predictors. All predictors retained their significance and the same correlation sign was observed in both data sets (see Appendix 5.3 and 5.4 for details).

Table 5.6 **Summary of Goodness of R^2 of the Regression Model for the Learning and Validation Data Sets in Urban and Rural Areas**

Data Set	Urban	Rural
Learning	0.7417	0.5801
Validation	0.7517	0.5762

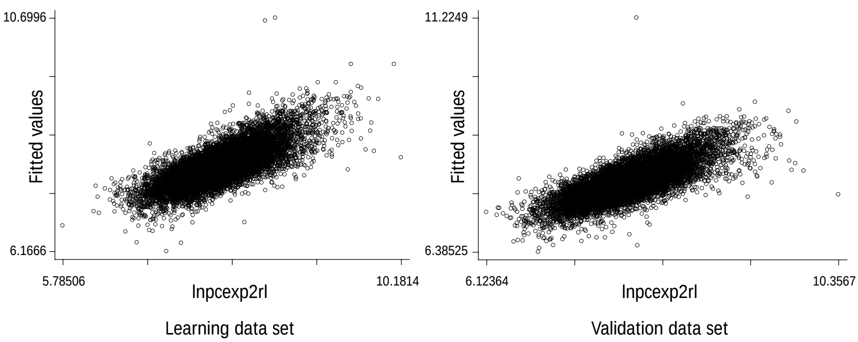
Source: Author's summary based on SUSENAS for the modeling development results.

Figure 5.3 **Residual Plot for the Rural Subsamples**



Note: This is to test homogeneity criteria of the residuals.
Source: Author's calculation based on 2002 VLSS.

Figure 5.4 **Actual Versus Predicted Values of Log Per Capita Expenditure for the Rural Subsamples**



lnpcexp2rl = natural logarithm of real per capita expenditure
Note: This is to test homogeneity criteria of the residuals.
Source: Author's calculation based on 2002 VLSS.

Diagnosing the models through a residual check, as shown in Figure 5.3, revealed that error variance is constant across observations for both rural subsamples, hence, the error term is homoscedastic. This is verified in Figure 5.4, which also proves linearity of the error.

The matched tabulation in Table 5.7 shows a good percentage match in the top and bottom quintiles, almost 60.0 percent for both. For the middle quintiles, the match is not very high, probably due to the small difference among adjacent households in terms of per capita expenditure. However, quintile 1 of the predicted log of PCE for the learning data set catches about 85.0 percent of total people in quintiles 1 and 2 of the actual values, that is, 59.6 percent and 25.4 percent, respectively. This is similar to the result in the validation data set. Therefore, if the purpose is to detect poor people and provide support, including people in quintile 1 of the predicted values can be relevant.

Table 5.7 **Matched Tabulation for the Rural Subsamples**

Learning Data Set		**Predicted Quintiles**					
		1	**2**	**3**	**4**	**5**	**Total**
Actual quintile	**1**	59.6	27.2	10.0	3.0	0.2	20.0
	2	25.4	32.8	25.6	13.7	2.5	20.0
	3	11.3	24.0	30.7	24.8	9.2	20.0
	4	3.1	12.6	24.4	34.3	25.4	20.0
	5	0.5	3.4	9.2	24.2	62.6	20.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0

Validation Data Set		**Predicted Quintiles**					
		1	**2**	**3**	**4**	**5**	**Total**
Actual quintile	**1**	59.8	26.7	10.8	2.5	0.3	20.0
	2	25.0	33.1	26.5	12.9	2.4	20.0
	3	10.5	23.6	30.1	27.3	8.5	20.0
	4	4.1	12.7	23.8	34.2	25.2	20.0
	5	0.6	3.9	8.7	23.1	63.7	20.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0

Source: Authors' calculation based on 2002 VLSS.

To further validate the models, mean values of the predicted log of PCE calculated from the two data sets were also compared. As shown in Table 5.8, the values of the two data sets are quite similar and show the stability of the model across the whole data set for rural areas.

Table 5.8 **Comparison of Mean Values of the Per Capita Expenditure for the Rural Subsample**

	Learning Data Set		***Validation Data Set***	
Quintile	*Actual Mean*	*Predicted Mean*	*Actual Mean*	*Predicted Mean*
1	1,321	1,557	1,326	1,552
2	1,926	2,066	1,925	2,067
3	2,441	2,447	2,422	2,446
4	3,138	2,941	3,142	2,941
5	5,091	4,342	5,090	4,310

Note: Total number of observations = 11,299
Source: Authors' calculation based on 1997/98 VLSS.

In Phase 2 for the rural areas, the model is applied to the 1997/98 VLSS, the results of which are presented in Tables 5.9 and 5.10 and Figures 5.5 and 5.6. As shown, almost all variables were still significant at 5 percent. Again, figures reveal that there was no heteroscedasticity in the error terms. This was an encouraging result given that the 1997/98 VLSS was conducted 4 years prior to the 2002 VHLSS.

At this point, the model now had 19 variables, including dummies, found to be very significant at the 5-percent level in the rural areas. There

Table 5.9 **Summary of Goodness of Fit of 1997/98 VLSS and Thanh Hao and Nghe An for Model Validation**

	Data Set	R-Squared
Subsample of VLSS 2002 and VLSS 1997/1998	Urban	0.6693
	Rural	0.5328
Survey in Thanh Hao and Nghe An	Learning	0.6039
	Validation	0.6100

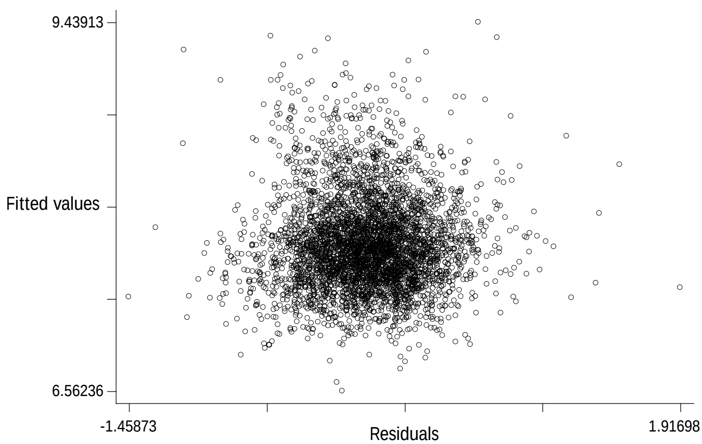
Source: Author's summary based on national and validation surveys.

Table 5.10 **Matched Tabulation for the Rural Subsamples Tested on the 1997/98 VLSS Rural Data Set**

		Predicted Quintile					
		1	2	3	4	5	Total
Actual Quintile	1	59.8	26.7	10.8	2.5	0.3	20.0
	2	25.0	33.1	26.5	12.9	2.4	20.0
	3	10.5	23.6	30.1	27.3	8.5	20.0
	4	4.1	12.7	23.8	34.2	25.2	20.0
	5	0.6	3.9	8.7	23.1	63.7	20.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0

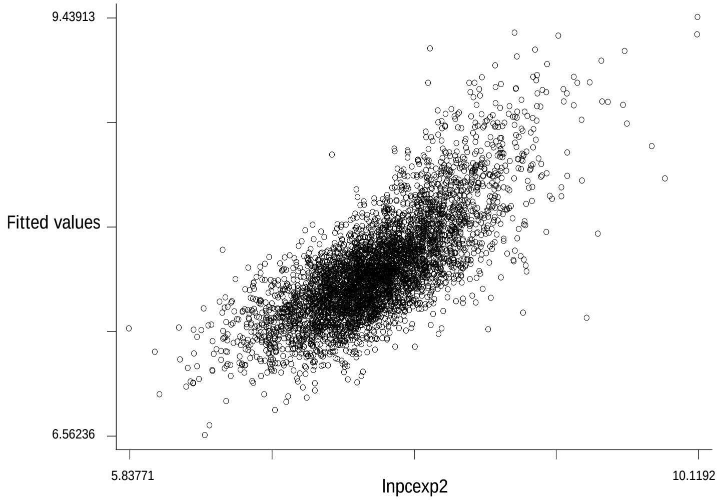
Source: Authors' calculations based on 1997/98 VLSS.

Figure 5.5 **Residual Plot for Rural Subsamples Tested on 1997/98 VLSS Rural Data Sets**



Note: This is to test homogeneity criteria of the residuals.
Source: Author's calculation based on 1997/98 VLSS.

Figure 5.6 **Actual Versus Predicted Values of Log Per Capita Expenditure for the Rural Subsamples Tested on 1997/98 VLSS Rural Data Sets**



lnpcexp2rl = natural logarithm of real per capita expenditure
Note: This is to test homogeneity criteria of the residuals.
Source: Author's calculation based on 1997/98 VLSS.

were 14 variables that belonged to five groups of household characteristics and 5 agricultural variables:

- Demographic: head's ethnicity, head's age, household size, marital status of the head, proportion of dependent people (aged <15 or >60 years)
- Assets: motorbike
- Housing: living area, electricity, toilet type, and house type
- Geographic: region
- Education: head's highest diploma, highest diploma of head's spouse, head's illiteracy
- Agricultural variables: agricultural land area, agricultural household, garden, rented-out land, proportion of members with main job in agriculture

This model was designed particularly for rural areas, therefore, variables relating to agricultural activities were of special concern. In this model, five agricultural variables are found to be significant in predicting household living standards. Households involved in agricultural activities in general have lower living standards than others, especially when there are more members involved in agriculture. However, if households were renting out agricultural land and maintained a garden at home, their living standards could improve significantly. Renting out agricultural land usually occurs when they have rights over a large piece of land or they have other higher income-earning activities.

The asset predictor (motorbike) has a positive relationship with the log of PCE.

Education, like in other studies, has a very strong effect on the living standards of households. The more education household heads have, the higher the household's living standards; and the less illiterate the heads are, the better the living conditions of the households.

The regional factor has strong impact. People living in the North Central Coast have lower living standards than people in other regions. This seems to be very reliable because these areas are always the hardest places to live in Viet Nam. The households in the South East area, including Ho Chi Minh City and the Mekong River Delta (the Rice Granary of Viet Nam), are better-off than in any other region, as shown by the very significant impact of the dummy variable for these regions.

The age of the household head has a positive impact on the household's living standards. The older the head, the better the living conditions. In addition, better household characteristics–that is, having a better toilet type, a larger living area, and access to electricity–means better living standards.

It is quite interesting that ethnic Kinh-Vietnamese and Chinese households have worse living standards than others. According to Dominique van de Walle and Dileni Gunewardena, this can be attributed to what they call as *quality gaps*, such as ethnic minorities receiving poor-quality education (Rama and Kim 2005).

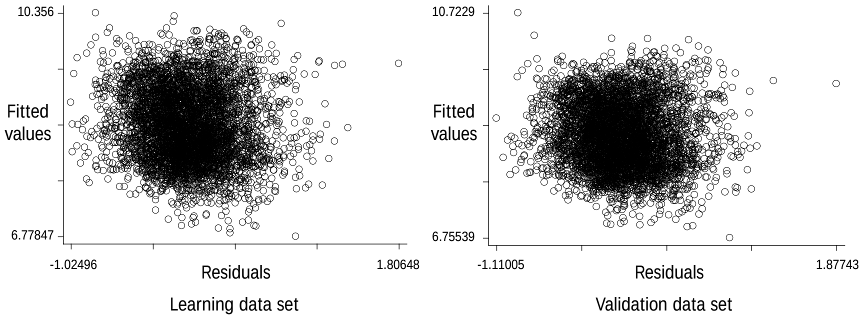
Households with more dependents and, especially, with more household members (larger household size) have lower living standards. Families living in semipermanent housing such as apartments and all temporary house-types also have lower living standards.

Urban Areas

The modeling process used for the rural data set was also applied to the urban data set and the model result was even better. As presented in Table 5.6, with only 3,455 observations for the learning data set and 3,454 in validation data set, the R-squared at 0.7417 and 0.7517, respectively, is higher for the urban data set than for the rural data set (see Appendix 5.7 and 5.8 for details). The assumption of homoscedasticity in the error term is also validated (Figures 5.7 and 5.8).

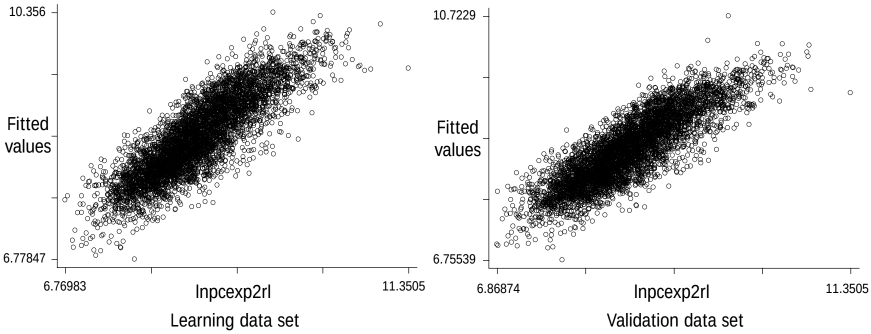
The matched tabulation in Table 5.11 also shows a good percentage match in the top and bottom quintiles, also almost 60 percent for both the learning and validation data sets. As it was for the rural areas, the match is not good for the middle quintiles.

Figure 5.7 **Residual Plot for the Urban Subsamples**



lnpcexp2rl = natural logarithm of real per capita expenditure
Note: This is to test homogeneity criteria of the residuals.
Source: Author's calculation based on 2002 VLSS.

Figure 5.8 **Log Per Capita Expenditure for Urban Subsamples—Actual Versus Predicted Values**



lnpcexp2rl = natural logarithm of real per capita expenditure
Note: This is to test homogeneity criteria of the residuals.
Source: Author's calculation based on 2002 VLSS.

As was done for the rural area subsamples, mean values of the predicted log of PCE calculated from the two data sets for the urban areas were compared to further validate the models. As exhibited in Table 5.12, the values of the two data sets are almost the same and reveal the stability of the model across the entire data set for urban areas.

With reference to Table 5.13 and Figures 5.9 and 5.10, testing results in Phase 2 for urban areas were also acceptable. As shown, almost all variables are still significant at 5 percent. Again, figures reveal that there is no heteroscedasticity in the error terms and the matched tabulation shows top and bottom quintiles are good matches.

Table 5.11 **Matched Tabulation for the Urban Subsamples on the 1997/98 VLSS Urban Data Set**

Learning Data Set		*Predicted Quintiles*					
		1	2	3	4	5	Total
Actual Quintiles	1	66.6	26.6	6.7	0.1	0.0	20.0
	2	24.6	44.1	25.9	5.4	0.0	20.0
	3	7.5	20.8	39.6	27.4	4.6	20.0
	4	1.2	7.4	23.6	42.0	25.9	20.0
	5	0.1	1.0	4.2	25.2	69.5	20.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0

Validation Data Set		*Predicted Quintiles*					
		1	2	3	4	5	Total
Actual Quintiles	1	67.0	27.1	5.2	0.7	0.0	20.0
	2	24.8	41.2	28.6	5.1	0.3	20.0
	3	6.4	24.0	39.6	25.3	4.6	20.0
	4	1.9	6.8	22.1	43.4	25.8	20.0
	5	0.0	0.9	4.3	25.5	69.3	20.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0

Source: Authors' calculation based on 2002 VLSS.

Table 5.12 **Comparison of Mean Values of Per Capita Expenditure for the Urban Subsamples**

	Learning Data Set		*Validation Data Set*	
Quintile	*Actual Mean*	*Predicted Mean*	*Actual Mean*	*Predicted Mean*
1	2,214	2,441	2,204	2,378
2	3,559	3,643	3,590	3,606
3	4,972	5,030	4,977	5,019
4	7,046	7,207	7,127	7,296
5	13,319	11,950	13,090	11,955

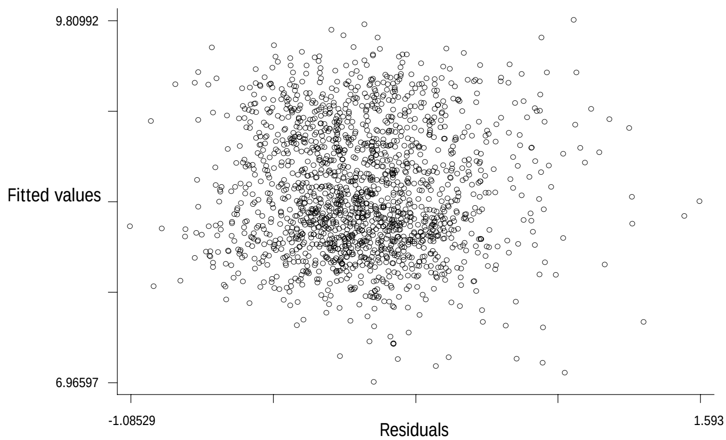
Note: Total number of observations = 3,454
Source: Authors' calculation based on 2002 VLSS.

Table 5.13 **Matched Tabulation for Urban Subsamples Tested on the 1997/98 VLSS Urban Data Set**

		Predicted Quintile					
		1	2	3	4	5	Total
Actual Quintile	1	65.0	26.3	8.7	0.0	0.0	20.0
	2	26.6	37.3	28.9	6.6	0.6	20.0
	3	6.4	27.8	35.0	25.4	5.5	20.0
	4	1.7	8.1	21.1	41.9	27.2	20.0
	5	0.3	0.6	6.4	26.0	66.8	20.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0

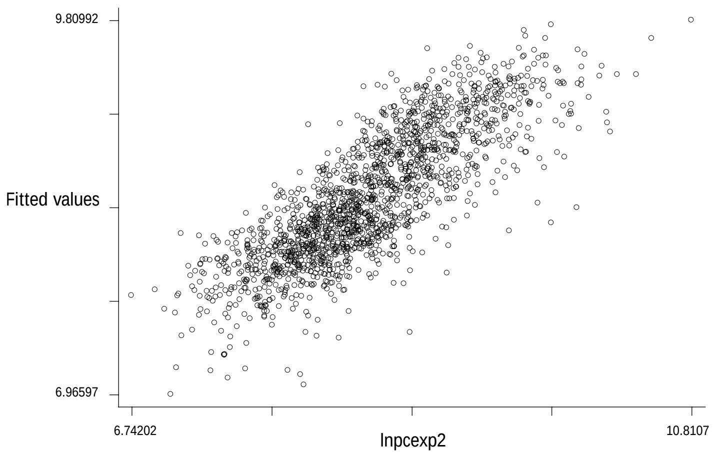
Source: Authors' calculation based on 1997/98 VLSS.

Figure 5.9 **Residual Plot of Urban Area Subsamples Tested on 1997/98 VLSS Urban Data Sets**



Note: This is to test homogeneity criteria of the residuals.
Source: Author's calculation based on 1997/98 VLSS.

Figure 5.10 **Log Per Capita Expenditure for the Urban Subsamples Tested on 1997/98 VLSS Urban Data Sets—Actual Versus Predicted Values**



Note: This is to test homogeneity criteria of the residuals.
Source: Author's calculation based on 1997/98 VLSS.

Some variables in the model for urban area subsamples tested in 1997/98 VLSS have the same signs of impact as in the rural areas. Households who have assets such as a gas cooker, motorbike, music mixer, refrigerator or freezer, rice cooker, or telephone are better-off. In addition, households are in better condition if the household head has had more education. If their house is relatively spacious and has a good toilet facility, then the family has good living conditions. Finally, those living in the South East have better living conditions than in other urban areas.

In contrast, households are poorer if household size is bigger and if there are more members of the family aged 15 years and below.

Results in Phase 3

From the modeling results of data sets for the provinces of Thanh Hoa and Nghe An (Table 5.9), R-squared values are found to be quite acceptable at 0.60 for the learning data set and 0.61 for the validation data set. For both data sets, at a 10-percent level of significance, all but one predictor (the proportion of members working in agriculture) are significant. The signs of correlations for models of both data sets are the same. Variables found significant were:

- Assets: colored TV, electric fan, motorbike, rice cooker, and water pump
- Demography: household size, proportion of household members less than 15 years old
- Education: head with college diploma or higher, spouse's educational attainment
- Employment: head's main occupation is white collar
- Housing: type of house and living area
- Health: number of household members hospitalized in the last 12 months

Ownership of a colored TV, electric fan, rice cooker, motorbike, or water pump dictates positive living standards in the two provinces. The same relationship is traced to the household head's educational attainment and main sectoral occupation (if a white collar job). In the subregion, a significant number of household heads in nonpoor households have white collar jobs. This may not be true for other areas, which may be why it was not significant in the model generated for the whole country.

Households with better house types–semipermanent or permanent–and larger houses also have better living conditions. Finally, the number of household members hospitalized in the past 12 months has a positive impact on living standards. It's possible that this means that members of poor households are seldom hospitalized because they don't have enough resources to pay for the hospitalization, and not because they seldom get sick.

As also discussed in previous results, household size and proportion of household members below 15 years old have negative relationships with living standards. In addition, the household experiences worse living conditions if the spouse of the household head has secondary educational attainment or below, or none at all. This may be attributed to less job opportunities in the subregion for people with these educational credentials (see Appendix 5.9–5.11 for details).

Results in Phase 4

An examination of the correlation between the different methods used for identifying poor households, shows that the correlation of poverty classifications based on self-assessment and enumerator's and hamlet chief's opinion is quite high (Table 5.14). In contrast, the correlation coefficients between these methods and PPM is quite low, ranging from 0.38 to 0.44. The coefficients are all significant at the 5-percent level.

Table 5.14 **Correlation between Different Methods Used for Identifying Poor Households**

Methods Used for Identifying Poor Households	*Self-Assessment*	*Enumerator*	*Hamlet Chief*	*Poverty Predictor Model*
Self-Assessment	1			
Enumerator	0.80	1		
Hamlet Chief	0.73	0.87	1	
Poverty Predictor Model	0.41	0.44	0.38	1

Source: Authors' calculation based on PPM questionnaire.

Table 5.15 shows that through self-assessment, 140 of the total 500 households surveyed are classified as poor, while this figure for PPM is only 110 of the total 500 households surveyed, resulting in a higher poverty rate based on self-assessment. This is not surprising since self-assessed poverty is usually high as households tend to be pessimistic when comparing their economic status with neighbors that are well-off. In terms of mismatch, 19 percent of PPM nonpoor are classified by self-assessment as poor and a rather large 34 percent of PPM poor are classified by self-assessment as nonpoor. The relatively large difference between the estimates based on PPM and self-assessment is broadly consistent with findings of similar works, such as the *Viet Nam Development Report* 2004 (World Bank 2004), on different poverty classifications.

Table 5.16 compares the classification based on the PPM and those based on the enumerator's assessment. It can be shown that almost 12 percent of PPM nonpoor were classified as poor by the enumerator, while 40 percent of the PPM poor were classified nonpoor by the enumerator. The enumerator's assessment is closer to the PPM classification with only 95 mismatched

Table 5.15 Matched Tabulation Between PPM Result sand SA-Based Poverty Classification

PPM Classification			SA Poverty Classification		
			Nonpoor	*Poor*	*Total*
Nonpoor	*Mean*		81.24	18.76	100.00
	Standard Error (%)		(2.51)	(2.51)	
	Number of Observations		319	71	390
Poor	*Mean*		34.07	65.93	100.00
	Standard Error (%)		(6.13)	(6.13)	
	Number of Observations		41	69	110
Total	*Mean*		72.26	27.74	100.00
	Standard Error (%)		(2.57)	(2.57)	
	Number of Observations		360	140	500

PPM = poverty predictor model; SA = self-assessment
Source: Authors' calculation based on PPM questionnaire.

households, compared with 112 mismatched households between self-assessed and PPM classifications. In addition, PPM-based poverty classification is only higher by three poor households compared with those classified as poor by the enumerator.

Table 5.16 Matched Tabulation Between PPM Results and EA-Based Poverty Classification

PPM Classification		EA-Based Poverty Classification		
		Nonpoor	*Poor*	*Total*
Nonpoor	*Mean*	88.21	11.79	100
	Standard Error (%)	(2.07)	(2.07)	
	Number of Observations	344	46	390
Poor	*Mean*	40.51	59.49	100
	Standard Error (%)	(6.36)	(6.36)	
	Number of Observations	49	61	110
Total	*Mean*	79.13	20.87	100
	Standard Error (%)	(2.33)	(2.33)	
	Number of Observations	393	107	500

EA = enumerators assessment; PPM = poverty predictor model
Source: Authors' calculation based on PPM questionnaire.

Comparing the classifications based on PPM and the hamlet chief's assessments, it can be observed from Table 5.17 that more households were classified as poor by the PPM. Based on the PPM, 110 poor households were classified as poor compared with 86 assessed as poor households by the hamlet chiefs. There were 98 mismatched households between these two classifications.

Among the four methods of classification, self-assessment classified the most number of poor with a total of 140 households. As mentioned earlier, self-assessed poverty status usually results in higher estimates because of the tendency of households to be pessimistic, sometimes hoping that they will

Table 5.17 **Matched Tabulation Between PPM Results and HCA-Based Poverty Classification**

PPM Classification			HCA-Based Poverty Classification		
			Nonpoor	*Poor*	*Total*
	Nonpoor	*Mean*	89.76	10.24	100
		Standard Error (%)	(1.95)	(1.95)	
		Number of Observations	353	37	390
	Poor	*Mean*	52.71	47.29	100
		Standard Error (%)	(6.49)	(6.49)	
		Number of Observations	61	49	110
	Total	*Mean*	82.71	17.29	100
		Standard Error (%)	(2.18)	(2.18)	
		Number of Observations	414	86	500

PPM = Poverty Predictor Model; HCA = Hamlet's Chief's Assesment
Source: Authors' calculation based on PPM questionnaire.

benefit from interventions if they declare themselves poor. The relatively close intervals of results among the PPM-based, enumerator's assessment, and hamlet chief's assessment methods could probably be accounted for by the fact that the PPM classification was actually based on easy-to-collect and observable variables, which could also be the same variables used by the enumerators and hamlet chiefs in assessing the poverty status of a household.

Aside from these assessments, the effectiveness of PPM can also be gauged by comparing the classification of households in the 2002 and 2004 VHLSSs using the consumption-based classification, since this model was developed through the VHLSS. Table 5.18 presents the comparison generated from using the 2002 VHLSS with 609 households classified as poor in this subregion based on household consumption and only 484 households classified as poor in the PPM.

Table 5.18 **Matched Tabulation Between PPM Results and Consumption-Based Poverty Classification**

PPM Poverty Classification			HCA Consumption-Based Classification		
			Nonpoor	*Poor*	*Total*
	Nonpoor	*Mean*	79.2	20.8	70.2
		Standard Error (%)	0.019	0.019	
		Number of Observations	903	243	1,146
	Poor	*Mean*	25.1	74.9	29.8
		Standard Error (%)	0.031	0.031	
		Number of Observations	118	366	484
	Total	*Mean*	63.1	36.9	100
		Standard Error (%)	0.02	0.02	
		Number of Observations	1,021	609	1,630

PPM = Poverty Predictor Model; HCA = Hamlet's Chief's Assessment
Source: Authors' calculation based on PPM questionnaire and 2002 VLSS.

Given these results, there is probably a need to refine the PPM to understand the relatively large discrepancy between the number of households classified as poor based on the PPM and those based on consumption data, considering that the VHLSS was used in developing the PPM.

Conclusion

Given the well-known problems in collecting household income or consumption expenditure data, poverty predictor models have been developed in recent years based on household demographic and asset characteristics which are easy to collect but significantly correlated to poverty. These models could be used to identify the poor households for intervention programs. This paper develops poverty predictor models for rural and urban areas in Viet Nam using the 2002 VHLSS survey data. The models are then tested for consistency and stability with 1997/98 VLSS data. The method is also verified using data from two relatively poor provinces and also from a pilot survey that takes into account local perceptions, among other information.

Overall, the poverty predictor models perform in a robust manner across alternative data sets. The variables in the model cover a wide range of easily verifiable information that include assets, such as TVs and motorbikes, and demographic characteristics, such as dependents and number of earning members, education, and housing conditions. Cross tabulations of actual and predicted values reveal that the models capture about 60 percent of the bottom-quintile households classified in terms of per capita expenditure distribution. Performance with respect to poor households also turns out to be similar.

Appendix

Appendix 5.1 **List of Primary Variables Identified from 2002 Viet Nam Living Standard Survey**

Variable Name	Description	Variable Name	Description
Tinh	Province	hunemp	Head is unemployed?
Huyen	District	num_unemp	Number of unemployed people
Xa	Commune/Ward	Hilliter	Head is illiterate?
Diaban	EAs	Pilliter	Husband/Wife is illiterate?
Hoso	Household Identification	Hdip	Head's highest diploma
Livingarea	Living area	Pdip	Husband/Wife's highest diploma
Housetype	Type of house	Hethnic	Head's ethnicity
Ownership	Do you own this house?	num_dep	Number of dependent people (age < 15 and > 60)
Payrent	Do you have to pay for rent?	num_u15	Number of age under-15 people
Rentpayee	Pay rent to whom?	num_o15	Number of age over-15 people
Otherhouse	Do you have other houses?	num_o60	Number of age over-60 people
Mfrout	Do you get any money from renting out any houses?	num_o70	Number of age over-70 people
Newbhouse	Did you have any newly built house in the last 12 months?	num_labor	Number of people in labor age (15 < age < 60)
Wsource	Main drinking water sources	num_child	Number of head's children
Toilet	Type of toilet	Hhsize	Household size
Electric	Electricity	prop_dep	Dependent proportion
Qui	Quarter of 2002	prop_u15	Proportion of < 15 people
Motorbike	If household has a motorbike?	prop_o15	Proportion of ≥ 15 people
Waterpump	If household has a water pump?	prop_o60	Proportion of > 60 people
Telephone	If household has a telephone?	prop_o70	Proportion of > 70 people
Video	If household has a video?	prop_labor	Proportion of people in labor age (15–60)
Colortv	If household has a colored TV?	Hsex	Head's sex
Bwtivi	If household has a black and white TV?	Hage	Head's age
Musicmixer	If household has a music mixer?	hmarital	Head's marital status
Refee	If household has a refrigerator?	reg8	8 regions
Elecfan	If household has an electric fan?	urban02	Urban: 1, Rural: 2
Gascooker	If household has a gas cooker?	wt30	Household weight
Ricecooker	If household has a rice cooker?	Hhszwt30	Individual weight
Nonfarm	Household with nonfarm activities	hhexp2rl	2002 real total household expenditure
num_inpatient	Number of times an inpatient	pcexp2rl	2002 real per capita expenditure
Inpatient	Any inpatient time?	prop_illi	Proportion of age ≥ 15 people illiterate
Hjbowner	Head's job owner	prop_studmem	Proportion of people studying in the last 12 months
hocc02	Head's sectoral occupation	prop_unemp	Proportion of unemployed people in the total age ≥ 15 people
prop_agri	Proportion of age ≥ 15 economically active people working in agriculture	Agrihh	Agricultural household
num_agri	Number of people involved in agricultural activities	Agland_area	Total agricultural land
rentedout	Household with land rented out	rentedin	Household with land rented in
agriser	If household does agricultural services	Garden	If household has a garden
Cow	If household has a cow	Brdfacs	If household has breeding facilities
Grinder	If household has a grinder	Mill	If household has a rice milling machine
Workshop	If household has a workshop	rplucker	If household has a rice plucker
Pullinmach	If household has a pulling machine	Store	If household has a store
Trailer	If household has a trailer	Plough	If household has a plough

Source: Authors' summary based on 2002 VLSS.

Appendix 5.2 **List of Candidate Variables for Rural Subsamples**

Variable Name	Description	Variable Name	Description
Colortv	If household has a colored TV?	pdip_3	Husband/Wife with upper secondary diploma
Elecfan	If household has an electric fan?	pdip_4	Husband/Wife with technical worker diploma
electric_t	Electricity	pilliter_t	Husband/Wife is illiterate?
gascooker	If household has a gas cooker?	Prop_dep_t	Dependent proportion
hage_t	Head's age	Prop_illi_t	Proportion of age ≥ 15 people illiterate
hdip_0	Head with primary diploma	Refee	If household has a refrigerator?
hdip_1	Head with lower secondary diploma	reg8_1	Red River Delta
hdip_2	Head with upper secondary diploma	reg8_2	North East
hdip_3	Head with technical worker diploma	reg8_3	North West
hdip_4	Head with professional secondary school diploma	reg8_4	North Central Coast
hdip_5	Head with junior college diploma and higher	reg8_5	South Central Coast
hdip_6	Head with primary diploma	reg8_6	Central Highlands
hethnic	Head's ethnicity	reg8_7	South East
hhsize	Household size	reg8_8	Mekong River Delta
hilliter	Head is illiterate?	ricecooker	If household has a rice cooker?
hjbowner_t	Head's job owner	Telephone	If household has a telephone?
hocc02_1	Head's sectoral occupation: agriculture, forestry, fishery	toilet_1	Flush toilet with septic tank/sewage pipes
hocc02_2	Head's sectoral occupation: manufacturing	toilet_2	Suilabh toilet
hocc02_3	Head's sectoral occupation: sales services	toilet_3	Double vault compost latrine
hocc02_4	Head's sectoral occupation: white collar	toilet_4	Toilet directly over the water
hocc02_5	Head's sectoral occupation: others	toilet_5	Others
hocc02_6	Head's sectoral occupation: others not working	toilet_6	No toilet
housetype_1	House type is villa or permanent house/ apartment with private bath/kitchen/toilet	Video	If household has a video?
housetype_2	House type is permanent house/ apartment without private bath/kitchen/toilet	waterpump	If household has a water pump?
housetype_3	House type is semipermanent house/ apartment	Wsource_1	Individual tap
housetype_4	Temporary house and others	Wsource_2	Public tap
Livingarea	Living area	Wsource_3	Deep drill well with pump
Motorbike	If household has a motorbike?	Wsource_4	Hand dug well, constructed well
Nonfarm	Household with nonfarm activities	Wsource_5	Deep well
pdip_0	Husband/Wife with no diploma	Wsource_6	Rain water
pdip_1	Husband/Wife with primary diploma	Wsource_7	River, lake, pond
pdip_2	Husband/Wife with lower secondary diploma	wsource_8	Bought water (in tank, bottled or in a jar), filtered spring water, and others
prop_agri	Proportion of age ≥ 15 economically active people working in agriculture	Agrihh	Agricultural household
num_agri	Number of people involved in agricultural activities	lnagland_area	Natural logarithm of total agricultural land
rentedout	Household with land rented out	rentedin	Household with land rented in
agriser	If household does agricultural services	Garden	If household has garden
Cow	If household has a cow	Brdfacs	If household has a breeding facilities
Grinder	If household has a grinder	Mill	If household has a rice milling machine
Workshop	If household has a workshop	rplucker	If household has a rice plucker
Pullinmach	If household has a pulling machine	Store	If household has a store
Trailer	If household has a trailer	plough	If household has a plough

Source: Authors' summary based on 2002 VLSS.

Appendix 5.3 Regression Model for Learning Data Set of Rural Subsamples

Variable	Variable Description	Estimate	Sign	Pr>|t|
Dependent Variable				
ln(pcexp2rl)	Natural logarithm of real per capita expenditure per year (best for 2002)			
Independent Variables				
Agrihh (Control variable)	Household with agricultural activities? Yes=1, No=0	-0.078	-	0.000
Garden	Household has a garden? Yes=1, No=0	0.049	+	0.006
Mill	Household has a mill? Yes=1, No=0	0.087	+	0.014
Agriser	Household does any agricultural services? Yes=1, No=0	0.045	+	0.054
rentedout	Household rented out its land? Yes=1, No=0	0.042	+	0.000
prop_agri	Proportion of members with main job in agriculture	-0.132	-	0.000
livingarea	Living area (m^2)	0.001	+	0.000
motorbike	Household has motorbike? Yes=1, No=0	0.237	+	0.000
Hethnic	Ethnicity Vietnamese and Chinese: 1, others: 2	0.068	+	0.000
electric_t	Household has access to electricity?	0.088	+	0.000
Hilliter	Is the head illiterate?	-0.071	-	0.000
hdip_0	Head's highest diploma: no diploma	-0.140	-	0.000
hdip_1	Head's highest diploma: primary school	-0.107	-	0.000
hdip_2	Head's highest diploma: lower secondary school	-0.094	-	0.003
hdip_3	Head's highest diploma: upper secondary school	-0.069	-	0.000
housetype_2	House type is permanent house/apartment without private bath/kitchen/toilet	-0.182	-	0.000
housetype_3	House type is semi-permanent house/apartment	-0.258	-	0.000
housetype_4	Temporary house and others	-0.385	-	0.000
No partner (control variable)	No husband/wife (widow, single, divorced)	-0.143	-	0.000
pdip_0	Head's husband/wife highest diploma: no diploma	-0.127	-	0.000
pdip_1	Head's husband/wife highest diploma: primary school	-0.135	-	0.000
pdip_2	Head's husband/wife highest diploma: lower secondary school	-0.125	-	0.018
pdip_3	Head's husband/wife highest diploma: upper secondary school	-0.088	-	0.000
reg8_4	North Central Coast	-0.072	-	0.000
reg8_7	South East	0.250	+	0.000
reg8_8	Mekong River Delta	0.291	+	0.000
toilet_1	Flush toilet with septic tank/sewage pipes	0.282	+	0.000
toilet_2	Suilabh toilet	0.177	+	0.000
toilet_3	Double vault compost latrine	0.091	+	0.001
Wsource_1	Individual tap	0.112	+	0.000
prop_dep_t	Dependent proportion	-0.236	-	0.000
Hhsize	Household size	-0.092	-	0.000
hage_t	Head's age	0.181	+	0.000
lnagriland	Natural logarithm of agricultural land area	0.009		0.000
Intercept		7.894	+	0.000

Model Statistics

pweight: wt30; Strata: Tinh; PSU: Diaban; Number of obs = 11299; Number of strata = 61; Number of PSUs = 880; Population size = 6523233; F(27,364) = 170.410; Prob>F = 0.000; R-squared = 0.5801

Source: Authors' calculation.

Appendix 5.4 **Regression Model for Validation Data Set of Rural Subsamples**

Variable	Variable Description	Estimate	Sign	Pr>|t|
Dependent Variable				
ln(pcexp2rl)	Natural logarithm of real per capita expenditure per year (best for 2002)			
Independent Variables				
agrihh	Household with agricultural activities? Yes=1, No=0	-0.093	-	0.000
garden	Household has a garden? Yes=1, No=0	0.031	+	0.017
mill	Household has a mill? Yes=1, No=0	0.099	+	0.001
agriser	Household does any agricultural services? Yes=1, No=0	0.043	+	0.017
rentedout	Household rented out its land? Yes=1, No=0	0.041	+	0.048
prop_agri	Proportion of members with main job in agriculture	-0.107	-	0.000
livingarea	Living area (m^2)	0.001	+	0.022
motorbike	Household has motorbike? Yes=1, No=0	0.241	+	0.000
hethnic	Ethnicity Vietnamese and Chinese: 1, others: 2	0.104	+	0.000
electric_t	Household has access to electricity?	0.070	+	0.000
hilliter	Is the head illiterate?	-0.071	-	0.000
hdip_0	Head’s highest diploma: no diploma	-0.145	-	0.000
hdip_1	Head’s highest diploma: primary school	-0.098	-	0.000
hdip_2	Head’s highest diploma: lower secondary school	-0.089	-	0.000
hdip_3	Head’s highest diploma: upper secondary school	-0.050	-	0.037
housetype_2	House type is permanent house/apartment without private bath/kitchen/toilet	-0.135	-	0.000
housetype_3	House type is semi-permanent house/apartment	-0.208	-	0.000
housetype_4	Temporary house and others	-0.356	-	0.000
nopartner	No husband/wife (widow, single, divorced)	-0.183	-	0.000
pdip_0	Head’s husband/wife highest diploma: no diploma	-0.153	-	0.000
pdip_1	Head’s husband/wife highest diploma: primary school	-0.144	-	0.000
pdip_2	Head’s husband/wife highest diploma: lower secondary school	-0.155	-	0.000
pdip_3	Head’s husband/wife highest diploma: upper secondary school	-0.122	-	0.000
reg8_4	North Central Coast	-0.077	-	0.000
reg8_7	South East	0.218	+	0.000
reg8_8	Mekong River Delta	0.291	+	0.000
toilet_1	Flush toilet with septic tank/sewage pipes	0.285	+	0.000
toilet_2	Suilabh toilet	0.211	+	0.000
toilet_3	Double vault compost latrine	0.078	+	0.000
wsource_1	Individual tap	0.122	+	0.001
prop_dep_t	Dependent proportion	-0.232	-	0.000
hhsize	Household size	-0.088	-	0.000
hage_t	Head’s age	0.170	+	0.000
lnagriland	Natural logarithm of agricultural land area	0.011		0.000
Intercept		7.888	+	0.000

Model Statistics

pweight: wt30; Strata: tinh; PSU: diaban; Number of obs = 11301; Number of strata = 61; Number of PSUs = 882; Population size = 6566241; F(27,364) = 200.620; Prob>F = 0.000; R-squared = 0.5762

Source: Authors’ calculation.

Appendix 5.5 **Regression Model of 2002 VLSS for Rural Areas Tested on 1997/98 VLSS Rural Subsamples**

Variable	Variable Description	Estimate	Sign	Pr>|t|
Dependent Variable				
ln(pcexp2rl)	Natural logarithm of real per capita expenditure per year (best for 2002)			
Independent Variables				
Agrihh (control variable)	Household with agricultural activities? Yes=1, No=0	-0.068	-	0.000
Garden	Household has a garden? Yes=1, No=0	0.051	+	0.006
Mill	Household has a mill? Yes=1, No=0	0.087	+	0.231
Agriser	Household does any agricultural services? Yes=1, No=0	0.062	+	0.154
rentedout	Household rented out its land? Yes=1, No=0	0.072	+	0.000
prop_agri	Proportion of members with main job in agriculture	-0.102	-	0.000
livingarea	Living area (m^2)	0.060	+	0.000
motorbike	Household has motorbike? Yes=1, No=0	0.312	+	0.000
Hethnic	Ethnicity Vietnamese and Chinese: 1, others: 2	0.059	+	0.000
electric_t	Household has access to electricity?	0.092	+	0.001
Hilliter	Is the head illiterate?	-0.097	-	0.032
hdip_0	Head's highest diploma: no diploma	-0.140	-	0.000
hdip_1	Head's highest diploma: primary school	-0.107	-	0.000
hdip_2	Head's highest diploma: lower secondary school	-0.094	-	0.003
hdip_3	Head's highest diploma: upper secondary school	0.018	-	0.169
housetype_2	House type is permanent house/apartment without private bath/kitchen/toilet	0.125	-	0.462
housetype_3	House type is semi-permanent house/apartment	-0.158	-	0.014
housetype_4	Temporary house and others	-0.226	-	0.000
Nopartner (control variable)	No husband/wife (widow, single, divorced)	-0.285	-	0.000
pdip_0	Head's husband/wife highest diploma: no diploma	-0.038	-	0.004
pdip_1	Head's husband/wife highest diploma: primary school	-0.124	-	0.001
pdip_2	Head's husband/wife highest diploma: lower secondary school	-0.221	-	0.118
pdip_3	Head's husband/wife highest diploma: upper secondary school	0.088	-	0.609
reg8_4	North Central Coast	-0.002	-	0.876
reg8_7	South East	0.224	+	0.000
reg8_8	Mekong River Delta	0.279	+	0.000
toilet_1	Flush toilet with septic tank/sewage pipes	0.389	+	0.032
toilet_2	Suilabh toilet	0.107	+	0.000
toilet_3	Double vault compost latrine	0.001	+	0.001
Wsource_1	Individual tap	-0.041	+	0.652
prop_dep_t	Dependent proportion	-0.195	-	0.000
Hhsize	Household size	-0.153	-	0.000
hage_t	Head's age	0.151	+	0.000
lnagriland	Natural logarithm of agricultural land area	0.007		0.001
Intercept		7.785	+	0.000

Model Statistics

pweight: wt; Strata: Reg10; PSU: commune; Number of obs = 4265; Number of strata = 7; Number of PSUs = 136; Population size = 6566241; F(27,364) = 84.000; Prob>F = 0.000; R-squared = 0.5328

Source: Authors' calculation.

Appendix 5.6 **List of Candidate Variables for Urban Subsamples**

Variable Name	Description	Variable Name	Description
Bwtivi	If household has a black-and-white TV?	pdip_2	Husband/Wife with lower secondary diploma
Colortv	If household has a colored TV?	pdip_3	Husband/Wife with upper secondary diploma
Elecfan	If household has an electric fan?	pdip_4	Husband/Wife with technical worker diploma
Gascooker	If household has a gas cooker?	pdip_5	Husband/Wife with professional secondary school diploma
hdip_0	Head with no diploma	pdip_6	Husband/Wife with junior college diploma and higher
hdip_1	Head with primary diploma	prop_dep_t	Dependent proportion
hdip_2	Head with lower secondary diploma	prop_illi	Proportion of age ≥ 15 people illiterate
hdip_3	Head with upper secondary diploma	prop_labor	Proportion of people in labor age (15–60)
hdip_4	Head with technical worker diploma	prop_o15_t	Proportion of age ≥ 15 people
hdip_5	Head with professional secondary school diploma	prop_studmem_t	Proportion of people studying in the last 12 months
hdip_6	Head with junior college diploma and higher	prop_u15	Proportion of age < 15 people
Hethnic	Head's ethnicity	refee	If household has a refrigerator?
Hhsize	Household size	reg8_1	Red River Delta
Hilliter	Head is illiterate?	reg8_2	North East
Hjbowner_t	Head's job owner	reg8_3	North West
hmarital_t	Head's marital status	reg8_4	North Central Coast
hocc02_1	Head's sectoral occupation: agriculture, forestry, fishery	reg8_5	South Central Coast
hocc02_2	Head's sectoral occupation: manufacturing	reg8_6	Central Highlands
hocc02_3	Head's sectoral occupation: sales services	reg8_7	South East
hocc02_4	Head's sectoral occupation: white collar	reg8_8	Mekong River Delta
hocc02_5	Head's sectoral occupation: others	ricecooker	If household has a rice cooker?
hocc02_6	Head's sectoral occupation: others not working	telephone	If household has a telephone?
housetype_1	House type is villa or permanent house/apartment with private bath/kitchen/toilet	toilet_1	Flush toilet with septic tank/sewage pipes
housetype_2	House type is permanent house/apartment without private bath/kitchen/toilet	toilet_2	Suilabh toilet
housetype_3	House type is semipermanent house/apartment	toilet_3	Double vault compost latrine
housetype_4	Temporary house and others	toilet_4	Toilet directly over the water
hsex_t	Head's sex	toilet_5	Others
Livingarea	Living area	toilet_6	No toilet
mfrout_t	Do you get any money from renting out any houses?	video	If household has a video?
Motorbike	If household has a motorbike?	waterpump	If household has a water pump?
musicmixer	If household has a music mixer?	wsource_1	Individual tap
num_child	Number of head's children	wsource_2	Public tap
num_dep	Number of dependent people (age < 15 and > 60)	wsource_3	Deep-drill well with pump
num_labor	Number of people in labor age (15 < age < 60)	wsource_4	Hand dug well, constructed well
num_o15	Number of age over-15 people	wsource_5	Deep well
num_u15	Number of age under-15 people	wsource_6	Rain water
otherhouse_t	Do you have other houses?	wsource_7	River, lake, pond
pdip_0	Husband/Wife with no diploma	wsource_8	Bought water (in tank, bottled or in a jar), filtered spring water, and others
pdip_1	Husband/Wife with primary diploma		

Source: Authors' summary based on 1998 and 2002 VLSS.

Appendix 5.7 Regression Results for Learning Data Set of Urban Subsamples

Variable	Variable Description	Estimate	Sign	Pr>|t|
Dependent Variable				
ln(pcexp2rl)	Natural logarithm of real per capita expenditure (best for 2002)			
Independent Variables				
gascooker	Household has a gas cooker? Yes=1, No=0	0.048	+	0.062
hdip_6	Household head's highest diploma is junior college or higher.	0.135	+	0.000
hhsize	Household size	-0.103	-	0.000
hmarital_t	Household head is not married yet	0.143	+	0.007
housetype_1	House type is villa or permanent house/ apartment with private bath/kitchen/toilet	0.259	+	0.000
housetype_4	No house, temporary, or other house types	-0.152	-	0.000
livingarea	Living area	0.002	+	0.000
motorbike	Household has a motorbike? Yes=1, No=0	0.180	+	0.000
musicmixer	Household has a music-mixer? Yes=1, No=0	0.091	+	0.000
num_u15	Number of age under-15 people in the household	-0.069	-	0.000
refee	Household has a refrigerator/freezer? Yes=1, No=0	0.181	+	0.000
reg8_4	North Central Coast	-0.205	-	0.000
reg8_6	Central Highland	-0.108	-	0.011
reg8_7	South East	0.296	+	0.000
ricecooker	Household has a rice cooker? Yes=1, No=0	0.100	+	0.000
telephone	Household has a telephone? Yes=1, No=0	0.146	+	0.000
toilet_1	Flush toilet with septic tank/sewage pipes	0.151	+	0.000
toilet_5	Other types of toilet	-0.087	-	0.012
wsource_1	Private tap	0.152	+	0.000
wsource_4	Constructed well	-0.064	-	0.021
wsource_5	Simple soiled well	-0.158	-	0.001
Intercept		8.432	+	0.000

Model Statistics

pweight: wt30; Strata: tinh; PSU: diaban; Number of obs = 3,455; Number of strata = 61; Number of PSUs = 443; Population size = 2,055,589; F(27,364) = 143.27; Prob>F = 0.0000; R-squared = 0.7417

Source: Authors' calculation based on 2002 VLSS.

Appendix 5.8 **Regression Results for Validation Data Set of Urban Subsamples**

Variable	Variable Description	Estimate	Sign	Pr>|t|
Dependent Variable				
ln(pcexp2rl)	Natural logarithm of real per capita expenditure (best for 2002)			
Independent Variables				
gascooker	Household has a gas cooker? Yes=1, No=0	0.113	+	0.000
hdip_6	Household head's highest diploma is junior college or higher	0.152	+	0.000
hhsize	Household size	-0.092	-	0.000
hmarital_t	Household head is not married yet	0.198	+	0.000
housetype_1	House type is villa or permanent house/ apartment with private bath/kitchen/toilet	0.223	+	0.000
housetype_4	No house, temporary, or other house types	-0.185	-	0.000
livingarea_t	Living area	0.002	+	0.000
motorbike	Household has a motorbike? Yes=1, No=0	0.152	+	0.000
musicmixer	Household has a music mixer? Yes=1, No=0	0.159	+	0.000
num_u15	Number of age under-15 people in the household	-0.072	-	0.000
refee	Household has a refrigerator/freezer? Yes=1, No=0	0.141	+	0.000
reg8_4	North Central Coast	-0.132	-	0.000
reg8_6	Central Highland	-0.111	-	0.007
reg8_7	South East	0.312	+	0.000
ricecooker	Household has a rice cooker? Yes=1, No=0	0.093	+	0.000
telephone	Household has a telephone? Yes=1, No=0	0.156	+	0.000
toilet_1	Flush toilet with septic tank/sewage pipes	0.163	+	0.000
toilet_5	Other types of toilet	-0.097	-	0.003
wsource_1	Private tap	0.121	+	0.000
wsource_4	Constructed well	-0.103	-	0.001
wsource_5	Simple soiled well	-0.164	-	0.001
Intercept		8.395	+	0.000

Model Statistics

pweight: wt30; Strata: tinh; PSU: diaban; Number of obs = 3,454; Number of strata = 61; Number of PSUs = 445; Population size = 2,126,854; F(27,364) = 156.52; Prob>F = 0.0000; R-squared = 0.7517

Source: Authors' calculation based on 2002 VLSS.

Appendix 5.9 **Regression Results of 2002 VLSS for Urban Areas Tested on 1997/98 VLSS Urban Subsamples**

Variable	Variable Description	Estimate	Sign	Pr>|t|
Dependent Variable				
ln(pcexp2rl)	Natural logarithm of real per capita expenditure (best for 2002)			
Independent Variables				
gascooker	Household has a gas cooker? Yes=1, No=0	0.103	+	0.001
hdip_6	Household head's highest diploma is junior college or higher	0.077	+	0.006
hhsize	Household size	-0.096	-	0.000
hmarital_t	Household head is not married yet.	0.082	+	0.136
housetype_1	House type is villa or permanent house/ apartment with private bath/kitchen/toilet	0.009	+	0.799
housetype_4	No house, temporary or other house types	-0.060	-	0.082
livingarea_t	Living area	0.001	+	0.004
motorbike	Household has a motorbike? Yes=1, No=0	0.321	+	0.000
musicmixer	Household has a music mixer? Yes=1, No=0	0.177	+	0.000
num_u15	Number of age under-15 people in the household	-0.031	-	0.004
refee	Household has a refrigerator/freezer? Yes=1, No=0	0.178	+	0.000
reg8_4	North Central Coast	-0.046	-	0.277
reg8_6	Central Highland	0.183603	dropped	0.000
reg8_7	South East	0.143	+	0.000
ricecooker	Household has a rice cooker? Yes=1, No=0	0.167	+	0.000
telephone	Household has a telephone? Yes=1, No=0	0.110	+	0.000
toilet_1	Flush toilet with septic tank/sewage pipes	0.224	+	0.000
toilet_5	Other types of toilet	0.085	+	0.014
wsource_1	Private tap	-0.049	-	0.223
wsource_4	Constructed well	-0.099	-	0.118
wsource_5	Simple soiled well	-0.111	-	0.080
Intercept		8.341	+	0.000

Model Statistics

pweight: wt; Strata: reg10; PSU: commune; Number of obs = 1,730; Number of strata = 3; Number of PSUs = 58; Population size = 3,878,496; F(27,364) = 110.72; Prob>F = 0.0000; R-squared = 0.6693

Source: Authors' calculation based on 1997/98 and 2002 VLSS.

Appendix 5.10 **Regression Results for Learning Data Set for Thanh Hao and Nghe An**

Variable	Variable Description	Estimate	Sign	Pr>|t|
Dependent Variable				
ln(pcexp2rl)	Natural logarithm of real per capita expenditure (best for 2002)			
Independent Variables				
colortv	Household has a colored TV? Yes=1, No=0	0.104	+	0.002
elecfan	Household has an electric fan? Yes=1, No=0	0.084	+	0.006
hdip6	Head with college diploma and up	0.144	+	0.074
hhsize	Household size	-0.086	-	0.000
hocc024	Head's main sectoral occupation: white collar	0.159	+	0.016
housetype_1	Villa or permanent house/apartment with private bath/kitchen/toilet	0.489	+	0.000
housetype_2	Permanent house/apartment without private bath/kitchen/toilet	0.158	+	0.001
housetype_3	Semipermanent house/apartment	0.129	+	0.001
livingarea	Living area (m^2)	0.002	+	0.000
motorbike	Household has a motorbike? Yes=1, No=0	0.244	+	0.000
num_inpatient	Number of household members who were in-hospital patients over the last 12 months	0.078	+	0.005
pdip1	Head's husband/wife with no diploma	-0.149	-	0.004
pdip2	Head's husband/wife with primary diploma	-0.151	-	0.005
pdip3	Head's husband/wife with lower secondary diploma	-0.098	-	0.014
prop_agri	Proportion of members working in agriculture	-0.043	-	0.439
prop_u15	Proportion of household members under 15 years	-0.256	-	0.000
ricecooker	Household has a rice cooker? Yes=1, No=0	0.123	+	0.000
waterpump	Household has a water pump? Yes=1, No=0	0.072	+	0.068
Intercept		7.820	+	0.000

Model Statistics

pweight: wt30; Strata: Tinh; PSU: Diaban; Number of obs = 705; Number of strata = 2; Number of PSUs = 39; Population size = 631,215.9; F(27,364) = 89.76; Prob>F = 0.0000; R-squared = 0.6039

Source: Derived from poverty predictor model validation questionnaire.

Appendix 5.11 **Regression Results for Validation Data Set for Thanh Hao and Nghe An**

Variable	Variable Description	Estimate	Sign	Pr>|t|
Dependent Variable				
ln(pcexp2rl)	Natural logarithm of real per capita expenditure (best for 2002)			
Independent Variables				
colortv	Household has a colored TV? Yes=1, No=0	0.085	+	0.001
elecfan	Household has an electric fan? Yes=1, No=0	0.111	+	0.006
hdip6	Head with college diploma and up	0.120	+	0.016
hhsize	Household size	-0.089	-	0.000
hocc024	Head's main sectoral occupation: white collar	0.160	+	0.046
housetype_1	Villa or permanent house/apartment with private bath/kitchen/toilet	0.383	+	0.000
housetype_2	Permanent house/apartment without private bath/kitchen/toilet	0.264	+	0.000
housetype_3	Semipermanent house/apartment	0.199	+	0.000
livingarea	Living area (m^2)	0.001	+	0.002
motorbike	Household has a motorbike? Yes=1, No=0	0.276	+	0.000
num_inpatient	Number of household members who were in-hospital patients over the last 12 months	0.093	+	0.000
pdip1	Head's husband/wife with no diploma	-0.100	-	0.032
pdip2	Head's husband/wife with primary diploma	-0.118	-	0.014
pdip3	Head's husband/wife with lower secondary diploma	-0.097	-	0.014
prop_agri	Proportion of members working in agriculture	-0.049	-	0.304
prop_u15	Proportion of household members under 15 years	-0.345	-	0.000
ricecooker	Household has a rice cooker? Yes=1, No=0	0.077	+	0.000
waterpump	Household has a water pump? Yes=1, No=0	0.067	+	0.036
Intercept		7.825	+	0.000

Model Statistics

pweight: wt30; Strata: Tinh; PSU: Diaban; Number of obs = 705; Number of strata = 2; Number of PSUs = 39; Population size = 641,897.7; F(27,364) = 113.25; Prob>F = 0.0000; R-squared = 0.61

Source: Derived from poverty predictor model validation questionnaire.

CHAPTER 6

Poverty Mapping and GIS Application in Indonesia: How Low Can We Go?

Uzair Suhaimi , Guntur Sugiyarto, Eric B. Suan, and Mary Ann Magtulis

Introduction

The overarching goal of the Asian Development Bank (ADB) is to reduce poverty, which is in line with Millennium Development Goal (MDG) No. 1 of halving poverty incidence by 2015. In this context, a systematic technique for identifying poor regions is very important in improving poverty reduction programs.

Most poverty indicators developed with national household survey data, however, are reliable only at very aggregated levels such as province or state, with a possibility of further disaggregation into urban and rural. Poverty indicators in Indonesia derived from the National Socioeconomic Survey (SUSENAS), for instance, are reliable only up to the provincial level by urban and rural areas. This level of aggregation may not be appropriate for various poverty reduction projects or programs. Therefore, the availability of poverty indicators at a more disaggregated geographical area is very essential, especially in the context of poverty targeting and other poverty reduction programs.

One way to develop poverty indicators for smaller areas is to use poverty mapping, which has been implemented in Indonesia since 1990 (Suryahadi and Sumarto 2003b). The main goal of poverty mapping is to generate reliable estimates of poverty indicators at disaggregated levels to better understand local specificities. It would otherwise not be possible to obtain such disaggregated indicators given the existing household survey data.

Poverty mapping results have been increasingly used to geographically target scarce resources (Baschieri and Falkingham 2005). Mapping results may also include other welfare indicators such as the health and nutritional status of the population. Box 6.1 highlights the benefits that poverty mapping can substantiate in policies, while, to present a balance view, Box 6.2 cites different concerns underlying the efficiency of the estimates from poverty mapping.

Box 6.1 **The Benefits of Mapping Poverty Indicators**

Poverty mapping is a method to estimate poverty indicators for more disaggregated geographic units that the household survey can not produce. With poverty mapping, poverty impact assessments can be conducted at more disaggregated levels. Results of poverty mapping can help define poverty, describe the situation and problem, identify and select interventions, and guide resource allocation. Geographically disaggregated data from these assessments can then be displayed in a map. Henniger (1998) pointed out that linking poverty assessments to maps provides new benefits such as:

- Poverty maps make it easier to integrate data from various sources and from different disciplines to help define and describe poverty.
- A spatial framework allows switching to new units of analysis, such as from administrative to ecological boundaries, and access new variables not collected in the original survey like community characteristics.
- Identifying spatial patterns with poverty maps can provide new insights into the causes of poverty. An example is how much of the physical isolation and poor agroecological endowments impediments are needed to escape poverty that affects the type of interventions to consider.
- The allocation of resources can be improved. Poverty maps can assist in deciding where and how to target antipoverty programs. Geographic targeting, as opposed to across-the-board subsidies, has been shown to be effective at maximizing the coverage of the poor while minimizing leakage to the nonpoor (Baker and Grosh 1994).
- With appropriate scale and robust poverty indicators, poverty maps can assist in the implementation of poverty reduction programs such as providing subsidies in poor communities and cost recovery in less poor areas.
- Poverty maps with high resolution can support efforts to decentralize and localize decision making.

Maps are powerful tools for visualizing spatial relationships and can be used very effectively to reach policy makers. They provide an additional return on investments in survey data, which often remain unused and unanalyzed after the initial report or study is completed.

Source: Author's summary.

The term *poverty mapping* has been used interchangeably to refer to an econometric modeling technique, or to generating a map of existing poverty indicators, or a combination of the two—estimating the poverty indicators and then generating their maps. Poverty mapping in this study refers to the last point meaning, i.e., poverty mapping modeling and developing a geographic information system (GIS) map application of the poverty mapping modeling results.

Box 6.2 **Some Recent Concerns on Poverty Mapping**

Poverty estimates from household income or expenditure surveys are normally available at the national or provincial level. To fill an obvious data gap in dealing with poverty issues in small areas like districts, subdistricts, and villages; Elbers, Lanjouw, and Lanjouw (2003a), introduced a poverty mapping technique which has been applied in several countries. This technique estimates correlates of poverty for a set of variables which are common to household surveys and censuses and then predicts poverty for smaller areas using census data.

In 2006, an independent committee evaluating the World Bank's research (http://www.worldbank.org/poverty/) raised some concerns about the precision of smaller-area poverty estimates of poverty mapping. In particular, the committee was concerned that the prediction errors in census blocks across space within a local area, say wards within a city or districts within a province, would not be independent, giving rise to spatial correlation in error terms. In the absence of reliable estimates, the committee thinks poverty maps would be of "limited usefulness." In view of this problem, poverty maps may be viewed as indicative rather than firm measures of the extent of poverty in small areas and should be used with other available indicators of poverty for decision-making processes.

Source: Author's summary.

Poverty mapping modeling based on data sets from household survey and census data reveals relationships between poverty and some variables common to both types of data sources. The modeling relationship is then applied to population census data to get estimates of poverty indicators of wider geographical areas. Finally, poverty maps are developed to achieve the following purposes:

- Develop more accurate and cost-effective targeting and monitoring of poverty reduction projects and programs.
- Improve ex-ante impact assessment of proposed projects and policies.
- Improve poverty analysis and statistical capacity.
- Foster good governance by increasing the transparency of government resource allocation and disseminating information about the geographic distribution of poverty to stakeholders.

Applications of Poverty Mapping Across Countries

Elbers, Lanjouw, and Lanjouw (2002, 2003a, 2003b, 2004) developed the technique of poverty mapping to use detailed information about living standards available in household surveys and wider coverage of censuses to estimate poverty indicators at relatively small areas. By combining the strengths of each source and the technique, the estimators can be used at a remarkably disaggregated level to create effective poverty maps for clusters of subregional levels.

Poverty mapping has been implemented successfully in a number of countries to generate disaggregated poverty indicators, as summarized in Table 6.1. A similar procedure was also applied by Arellano and Meghir (1992) in a labor supply model using the United Kingdom's Family Expenditure Survey to estimate models of wages and other income conditioning on variables common across two samples.

Table 6.1 **Applications of Poverty Mapping in Some Selected Countries**

Country/ Reference	*Focus of Estimation*	*Lowest Disaggregation Level*
Cambodia Fujii, T. (2005)	Child Malnutrition Indicators	Commune
Ecuador Hentschel et al. (2000)	Basic needs and welfare indicators	Parish (lowest administrative area)
Indonesia SMERU (2005)	Poverty incidence	Village
Madagascar Mistiaen et al. (2001)	Welfare indicators	Commune (lowest administrative area)
Mozambique Simler and Nhate (2003)	Welfare, poverty (incidence and gap) and inequality measures	Village
Philippines World Bank (2005)	Poverty incidence, gap and severity	Municipality (urban and rural)
South Africa Alderman et al. (2002)	Poverty incidence	Magisterial district and transitional local council
Tajikistan Baschieri and Falkingham (2005)	Poverty incidence based on estimated consumption expenditure and food consumption expenditure	Rayon (district) and Jamoat (lowest administrative area)
Viet Nam Minot (1998)	Household characteristics as poverty indicators	District

Source: Authors' compilation.

Demombynes et al. (2001) constructed estimates of local welfare for many countries, while Henstchel et al. (2000) demonstrated how sample survey data can be combined with census data to yield predicted poverty rates for the population covered by the census. The use of geographic poverty maps was explored by Mistiaen et al. (2002) in Madagascar by combining detailed information from the household survey with the population census, replicating the method used by Elbers, Lanjouw, and Lanjouw (ELL Method). Cluster estimation was also used by Fujii (2005) to conduct small-area estimations of child nutrition status using the Cambodia Demographic and Health Survey. In his study, he extended the ELL model by identifying two layers of specific structure of error terms unique to nutrition indicators.

Poverty mapping studies for generating disaggregated welfare indicators have some similarities. The methodology is an extension of small-area estimation (Ghosh and Rao 1994, Rao 1999), i.e., applying the developed estimators based on small surveys to population census characteristics. Box 6.3 summarizes poverty mapping conducted for Pakistan, where the number of poor is estimated at the district level through poverty predictor modeling.

Box 6.3 **Poverty Mapping for Pakistan**

There are different ways to implement poverty mapping. One method is to produce maps of available poverty indicators and some relevant household characteristics (e.g., education, health, and other demographic information) directly from existing administrative or household survey data. Another method is to first estimate the number of poor households at the lowest possible disaggregated level, i.e., at district, subdistrict or village, through poverty modeling and then map out the result. This second method is done by using household characteristics available from survey and census data sets. Finally, a third method is to combine the first two methods by mapping poverty indicators from administrative or survey data as overlays on the map of poverty measures estimated through the model.

In poverty mapping done for Pakistan, the second approach was employed with an additional poverty incidence map using survey data with limited coverage. Two sets of thematic maps were also generated showing household characteristics by districts based on the 2001 Pakistan Socioeconomic Survey and the 1998 Population Census.

Three steps were involved in identifying poverty predictors and estimating poverty incidence at the district level. The first step was to use a multivariate regression model, where the dependent variable was per capita expenditure per month and the independent variables were various household characteristics. The next step was to use a probit model, where the dependent variable was poverty status, that is, a value of *1* is assigned if estimated per capita expenditure is below the poverty line, *0* if otherwise. This time the model estimation was done for every district. Based on both models, the poverty predictor variables found were household size, high dependency ratio, and low education. The final step was to implement multivariate poverty modeling using the estimated poverty incidence for every district as dependent variable and the significant predictors that resulted from the previous steps, but the data used were from the census. The result revealed estimated poverty incidence for 108 districts with the three most important predictors being family size, high dependency ratio, and education (Siddiqui 2005).

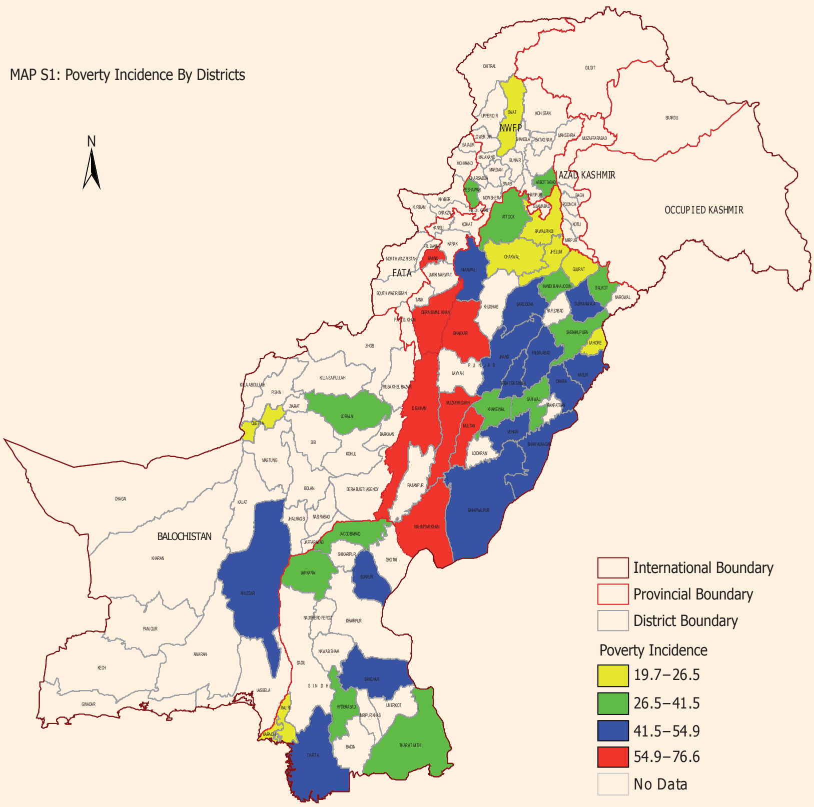
Figures 6.1 displays geographically referenced information on poverty incidence by district based on household survey data for only 71 districts in Pakistan. Figure 6.2 shows estimated poverty incidence based on poverty predictor modeling results for 108 districts in Pakistan. Figures 6.1 shows that incidence varies significantly across districts. The incidence of poverty is highest in Muzaffargarh (76.6 percent) and lowest in Panjgur (15.4 percent). Figure 6.2 reflects that poverty is not only concentrated in the southern part of Punjab but also in the central part of Balochistan and the upper part of the North Western Frontier Province.

continued on next page

Source: Nabeela 2005, ADB 2005b.

Box 6.3 continued

Figure 6.1 **A Poverty Map of Pakistan Showing Survey-Based Poverty Incidences**



Source: Based from the 2001 Pakistan Socio-Economic Survey.

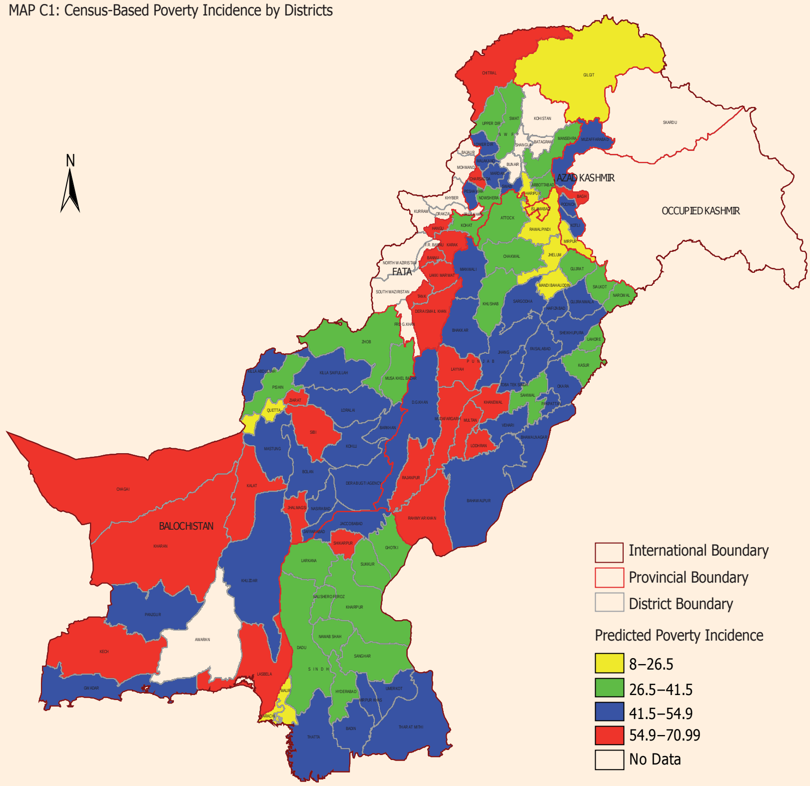
continued on next page

The construction of poverty maps for small administrative areas was also conducted in Indonesia as early as 1990. For allocating the poverty reduction fund as part of the Presidential Instruction on Disadvantaged Villages (IDT), entitled poor villages were identified based on a scoring system developed from a composite index of variables from the village census (Village Potential Statistics or *Potensi Desa*–Podes) data, complemented with the personal evaluation and perception of the subdistrict leader (*Camat*).

Box 6.3 continued

Figure 6.2 **A Poverty Map of Pakistan Showing Model-Based Poverty Incidences**

MAP C1: Census-Based Poverty Incidence by Districts



Source: Based from the 1998 Population Census of Pakistan.

The poverty mapping results identify possible causes of poverty, that suggest that geographically targeted policy measures may be used to alleviate poverty. The results can also be used for assessing the impact and effectiveness of poverty reduction programs.

Source: Nabeela 2005, ADB 2005b.

In another instance, the government's Family Welfare Development Program used a different classification system in defining the welfare status of families, i.e., according to some specific criteria such as religious practice, frequency of eating, pieces of clothing owned, types of house floor, and type of health services used. For a family to be classified as one with the highest welfare status, it has to satisfy a total of 24 indicators. Box 6.4 summarizes this welfare classification system.

Box 6.4 **Welfare Classification System of the Family Welfare Development Progam of Indonesia**

The Indonesian National Family Planning Movement has evolved from a fledgling program in the early 1970s into what it is now—a community and social development movement. From a purely clinical family planning approach, it has now become a comprehensive family development movement. The basis of its field operations is the annual family registration, undertaken January–March each year and based on 24 indicators. The hierarchical family welfare classification, or what is called the *family prosperity status*, is summarized below with the variables classified by stage of prosperity. It is important to emphasize that this registration is mainly for operational purposes, i.e., these variables serve as intervention points to elevate the prosperity status of each family.

This welfare classification system had also been used in the National Family Planning Coordinating Board's (BKKBN's) Family Prosperous Programme to improve family welfare (including family planning) autonomously after gaining a "prosperous family" status.

Source: Summarized from Weidemann (1998).

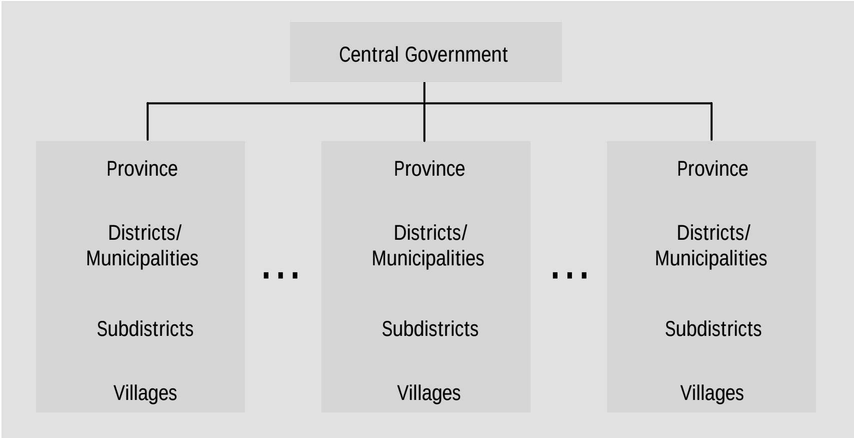
Moreover, an independent Indonesian institution for research and public policy studies, the Social Monitoring and Early Response Unit (SMERU), developed a tool for better targeting the poor by implementing poverty mapping. Using the ELL method, poverty indicators for small areas were estimated and GIS maps of the results were developed. The poverty mapping developed in this paper further refines the SMERU work by introducing some new features such as a dynamic "traffic-light" classification system that uses red, yellow, and green to represent high, moderate, and low poverty incidence; options for changing default cutoff points; and the option to overlay the poverty maps with graphs of variables taken from the Podes (which collects information on infrastructure and social facilities).

Study Background

Indonesia is the fourth most populous country and is the biggest archipelago (having the most number of islands) in the world. The first level of administration below the central government administration is the province. Each province is then further divided into districts (*Kabupaten*) or municipalities (*Kotamadya*), subdistricts (*Kecamatan*), and villages (*Desa/Kelurahan*) as the lowest administrative level (Figure 6.3).

Indonesia has relatively high poverty incidence compared with its neighbors like Malaysia and Thailand. In 2004, for instance, about 36 million people in Indonesia lived below the poverty line and the corresponding poverty incidences in total, rural, and urban areas were 16.7 percent, 20.3 percent, and

Figure 6.3 **Administrative Structures in Indonesia**



Source: Authors' summary.

13.5 percent, respectively. On the other hand, poverty incidence in Malaysia in 1999 was 7.5 percent and in Thailand in 2002 it was 9.9 percent.[1]

Poverty lines and poverty indicators in Indonesia were calculated using data from the SUSENAS, which collects among others, data on household income expenditures on different kinds of goods and services that can be used for calculating poverty indicators. The official poverty indicators were first published by Badan Pusat Statistik (BPS) Indonesia in 1984 for the period 1976–1984. Since then, poverty indicators have been estimated annually as part of the government program to reduce poverty. This program was intensified in 1994 with the implementation of the IDT program. Unfortunately, the economic crisis in 1997 resulted in an increase in the number of poor in Indonesia.

Table 6.2 shows poverty indicators in Indonesia from 1976 to 2003. Economic development was able to reduce poverty significantly in the early years. In 1976, 54 million people or 40 percent of the population were poor and the number was reduced to below 35 million or 22 percent in 1984, a remarkable reduction of almost 19 percentage points in a period of 8 years. The reduction slowed down in subsequent years as oil revenues declined. By 1993, 14 percent of the population was poor and in 1996 the headcount ratio was only 11.3 percent–the lowest in the history of the country. This trend was reversed drastically by the economic crisis in 1997, so much so that in 1998 the poverty incidence increased to 24 percent. From 1999, it has remained fairly constant at around 17 to 19 percent.

---

1 ADB Poverty and Development Indicators Database Online Query (http://lxapp1.asiandevbank.org:8030/sdbs/jsp/).

Table 6.2 **Poverty in Indonesia, 1976–2003**

Year	Poverty Line (Rp/capita/ month)		Headcount Ratio (%)			Poverty Incidence (million)		
	Urban	Rural	Urban	Rural	Total	Urban	Rural	Total
1976	4,522	2,849	38.8	40.4	40.1	10	44.2	54.2
1978	4,969	2,981	30.8	33.4	33.3	8.3	38.9	47.2
1980	6,381	4,449	29.0	28.4	28.6	9.5	32.8	42.3
1981	9,777	5,877	28.1	26.5	36.8	9.3	31.3	40.6
1984	13,731	7,746	23.1	21.2	21.6	9.3	25.7	35
1987	17,381	10,294	20.1	16.1	17.4	9.7	20.3	30
1990	20,614	13,295	16.8	14.3	15.1	9.4	17.8	27.2
1993	27,905	18,244	13.5	13.8	13.7	8.7	17.2	25.9
1996	38,426	27,413	9.7	12.3	11.3	7.2	15.3	22.5
1999	89,845	69,420	15.1	20.2	18.2	12.4	25.1	37.5
2000	91,632	73648	14.6	22.4	19.1	12.3	26.4	38.7
2001	100,011	80,382	9.8	24.8	18.4	8.6	29.3	37.9
2002	130,499	96,512	14.5	21.1	18.2	13.3	25.1	38.4
2003	138,803	105,888	13.6	20.2	17.4	12.2	25.1	37.3

Rp = rupiah
Source: Sugiyarto, Oey-Gardiner, and Triaswati (2006).

The calculation of poverty indicators in Indonesia is based on the official poverty line, which is estimated at the provincial level with different poverty lines for urban and rural areas. The poverty lines have been estimated as the cost of consuming a food commodity basket of 2,100 calories per capita per day and some essential nonfood items for a given reference population.

Poverty incidence in Indonesia is widely dispersed across regions and provinces. For instance, poverty incidence varied from 3.4 percent in the province of Jakarta to 41.8 percent in Papua. Therefore, information on where the poor people are located is important, but such information is severely constrained by the design of the SUSENAS. Although the survey is conducted every year, its limited sample size and distribution only allow for the calculation of poverty indicators down to the provincial urban and rural levels.

To estimate poverty indicators at lower administrative levels, such as for district to village levels, poverty mapping was implemented using the 1999 SUSENAS, 2000 Population Census, and 2000 Podes. The results show that reliable poverty indicators can be generated at the subdistrict level with the standard errors of estimates at less than 10 percent. At the village level, however, the standard errors of the estimates increased at nearly 14 percent, making them less reliable. Detailed results of this poverty mapping are available from BPS Indonesia.

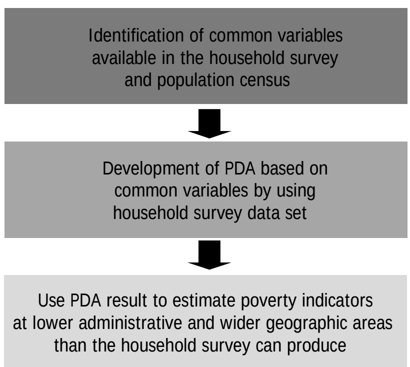
Modeling Developments

The methodology applied to this study, the ELL, is described in detail in Elbers, Lanjouw, and Lanjouw (2002, 2003a, and 2003b). The first major step in the application of the cluster method was running the regression models, based on the household per capita measure of consumption expenditure, on some exogenous variables found in both the household survey and population census. The household survey variables used in this poverty determinant analysis had to be strictly comparable to the variables in the population census.

The second major step was to estimate per capita consumption using the coefficients and residual terms randomly drawn from the estimated distribution as provided in the first step. The imputed consumption was, in turn, used to estimate poverty and inequality measures at the lowest administrative level, that is, the village level.[2] Simulation was done to arrive at robust point estimates with minimum standard error.[3]

Figure 6.4 shows the steps in implementing poverty mapping modeling. The common variables are identified according to some diagnostic tests in terms of relationships and distributional characteristics distinct to both the household survey and population census. Constrained to the underlying properties of the disturbance errors (idiosyncratic error), a cluster model is developed within the scope of poverty determinant analysis to identify

Figure 6.4 **Poverty Mapping Modeling**



PDA = poverty determinant analysis
Source: Authors' summary.

2 The process uses a computer program developed by Qinghua Zhao of the World Bank's Development Research Group (Qinghua 2002).

3 See Elbers, Lanjouw, and Lanjouw (2002, 2003a, and 2003b) for a more detailed description of the methodology.

significant parameters that would fit the census data. Finally, the parameter is subjected to a larger coverage area as depicted by the census data but bound by acceptable standard errors (model error and computational error).

Data Sources

Among the various surveys conducted by BPS Statistics Indonesia, the SUSENAS is the most appropriate data source for estimating poverty incidence due to the inclusion of consumption data. Besides the consumption data, the survey also covers numerous data items on population characteristics, such as demographic, education, health, employment, and housing characteristics which are also found in the population census. This study used the complete population census of 2000 for the purpose of providing the basic characteristics down to the lowest administrative levels, i.e., national, district, subdistrict, and village. In addition, accompanying every census is a Podes that collects information at the village levels. This information is intended to examine village potential in economic, social, and other aspects. Accordingly, other poverty-related indicators derived from the Podes can be overlaid with the poverty mapping results for spatial analysis.

Using the cluster-estimation method, poverty indices at the level of smaller administrative areas are estimated by combining the SUSENAS, Podes, and the complete 2000 Population Census data. Even though the SUSENAS is not designed to provide poverty estimation at levels lower than the province, it does supply consumption data that are required for estimating poverty measures. The census, on the other hand, does not cover consumption data but provides basic characteristics of individual households that make poverty estimation at the lowest level of administration possible.

In summary, poverty rate estimation as part of the poverty mapping is implemented using data sets from the following sources:

- SUSENAS Consumption Module (1999), which provides data on food and nonfood consumption. Total sample size of the survey is about 65,000 households throughout the country and is allocated proportionately in all provinces except Maluku, Maluku Utara, and Papua.
- SUSENAS Core (1999), which provides data on other individual and household characteristics and is used in implementing the cluster models. Total sample size is about 200,000 households and is allocated proportionately in all provinces except Maluku, Maluku Utara, and Papua.
- Population Census (2000), which provides data on individual and household characteristics. Data are used for simulation of various models for optimal estimation of poverty and inequality measures.

In addition, data generated are aggregated for the village level to produce community variables.

- Podes Census (2000), which provides community (i.e., village) data of approximately 69,000 villages. This is used to identify the so-called spatial distributional effects of poverty. The Podes covers all villages throughout the country and is used as the main data source to derive some geographic and background variables of poverty. The resulting characteristics are recommended for use as layers in poverty maps. In addition, the 2000 Master File of Villages (MFD) is used to link the four data sets. MFD is also employed to detect changes in villages during the period 1999–2000 to ensure the accuracy of village data.

Table 6.3 presents the determinants of poverty from each of the data sources. Using the common variables found in the census and survey data sets, and the variables that come from the Podes, consumption regression models were run to estimate the distribution of coefficients and residual terms. To provide more explanatory power for log per capita expenditure, the distribution and the summary statistics of each candidate variable were checked using Student t-statistics to compare data from the census and the survey. The variables with different distribution as shown in the summary statistics were excluded from the model. Checking for distribution and summary statistics is done at every stratum (province, urban and rural). Some variables used in determining the urban score for a village were composite indices. Table 6.4 lists the variables and their corresponding attributes and scores used in the construction of the urban score.

Table 6.3 **List of Variables Used in the Cluster Model Building in Indonesia**

Source	*Variable*
SUSENAS	Log expenditures per capita per month
SUSENAS/Podes/Census	Demographic Characteristics Education Occupation Health Infrastructure

SUSENAS = National Socioeconomic Survey: Podes = Village Potential Survey
Source: Authors' summary.

Table 6.4 **Variables Used in Constructing Urban Score**

Variable/Classification
1. Population density per km^2
2. Percentage of agricultural households
3. Percentage of households with electricity
4. Percentage of households with TVs
5. Accessibility to urban facilities
A. Kindergarten
B. Junior High School
C. Senior High School
D. Market with semi permanent or permanent building
E. Movie, theater/cinema
F. Shopping areas
G. Hospital
H. Hotel, billiards, amusement center
6. Village Total Score (5.A – 5.H)
7. Urban supporting facilities (only for urban)
A. Public lighting
B. Public bank
C. Public telephone/telecommunications shop
D. Supermarket/Department store
8. Total Score of Supporting Facility (7.A – 7.D)
9. Grand Total of Village Score (6 + 8)
10. Percentage of land area for other buildings other than housing

Source: Authors' summary.

In addition to common variables that satisfy the t-test, the interaction and higher-order variables (until the third order) derived from two or more well-tested single variables were also included. The cluster-estimation model is basically a prediction model and, hence, endogeneity problems are ignored.

In the prediction model, the dependent variable was the logarithm transformed per capita consumption as provided by the 1999 SUSENAS Consumption Module. The regression models were run for all provinces and, separately, for urban and rural areas.

Definitions and Properties of Estimators

The assimilation of individual characteristics from the SUSENAS and the 2000 Population Census was very similar to synthetic estimation used in small-area geographic modeling. The observed per capita household consumption in the SUSENAS was used as a function of a vector of variables characterized in both survey and census[4]:

$$\ell n\, y_{ch} = \mathrm{E}[\ell n\, y_{ch} \mid x_{ch}] + \mu_{ch} \tag{1}$$

where

y_{ch} : per capita consumption for household h and cluster c
x_{ch} : socio-economic characteristic of household h in cluster c
μ_{ch}- vector of disturbances

Using a linear approximation of the conditional expectation (Equation 1), the observed log per capita consumption expenditure can be expressed as follows:

$$\ell n(y_{ch}) = x_{ch}\,\beta + \mu_{ch} \quad (\textit{Beta}\text{ model}) \tag{2}$$

where β is a vector of c parameters and μ_{ch} is disturbance terms satisfying $E[\mu_h \mid \chi_h] = 0$.

By design, the SUSENAS does not provide spatial information. Therefore, the disturbance terms, as shown in Equation 2, include spatial effects and heteroskedasticity[5] to improve the model. The following formula is used for spatial effects:

---

4 Characteristics must have the same accuracy in the manner that definitions of each source are the same.

5 In the case of poverty mapping of Tajikistan (Baschieri and Falkingham 2005), heteroskedasticity appeared to be significant in some strata. In order to capture this, the alpha model was implemented only to result in a low R-squared. Hence, the heteroskedasticity component was not estimated; instead, a location component was estimated where possible.

$$\mu_{ch} = \eta_c + \varepsilon_{ch} \tag{3}$$

Here, η_c is a *cluster* component and ε_{ch} is a household component. On the average at village level, distribution terms can be expressed as follows:

$$\mu_{c.} = \eta_c + \varepsilon_{c.} \tag{4}$$

and then,

$$\mathrm{E}[\mu_c^{\ 2}] = \sigma_\eta^{\ 2} + \mathrm{var}(\varepsilon_{c.})$$

$$= \sigma_\eta^{\ 2} + \tau_c^{\ 2}$$

In the above equation, η_c and ε_{ch} are assumed to be normally distributed and independent from each other. Following Elbers, Lanjouw, and Lanjouw (2002), the estimated variance of spatial effects can be expressed as follows:

$$\mathrm{var}(\hat{\sigma}_\eta^{\ 2}) = \sum_c [a_c^{\ 2}\,\mathrm{var}(\mu_{c.}^{\ 2}) + b_c^{\ 2}\,\mathrm{var}(\hat{\tau}_c^{\ 2})] \tag{5}$$

In the absence of spatial effect, η_c, equation 3 becomes simpler, $\mu_{ch} = +\varepsilon_{ch}$.

However, this is normally an unrealistic assumption. Following Elbers, Lanjouw, and Lanjouw (2002), the residual can be explained by a logistic model that regresses the transformed ε_{ch} with household characteristics:

$$\ell n\left[\frac{\varepsilon^2{}_{ch}}{A - \varepsilon^{\ 2}{}_{ch}}\right] = Z_{ch}^T\hat{\alpha} + r_{ch} \quad (\textit{Alpha}\ \text{model}) \tag{6}$$

Here, A is set as $A = 1.05 * \max\left\{\varepsilon_{ch}^{\ 2}\right\}$, and r is a residual.

Estimated variance of ε_{ch} can be calculated using the following equation:

$$\hat{\sigma}^{\ 2}{}_{\varepsilon,ch} = \left[\frac{AB}{1+B}\right] + \frac{1}{2}\hat{V}ar(r)\left[\frac{AB(1-B)}{(1+B)^3}\right] \tag{7}$$

Here $B = \exp\{Z_{ch}^T\hat{\alpha}\}$

Equation 7 suggests the generalized least squares model is employed in Equation 2 instead of the ordinary least squares model.

In Equation 2, per capita logarithmic consumption $\ell n(y_{ch})$ as provided by the 1999 SUSENAS Consumption Module serves as the dependent variable. For explanatory variables x_{ch} all common variables found in both the 1999 SUSENAS Core and 2000 population data sets (both L1 and L2 schedules) can serve as candidate variables to be included in the model.

Properties considered:

- Presence of disturbance error at households' consumption expenditure from their expected value (μ_{ch}). This is proportional to the size of the population of households.
- Variance in the first-stage estimate of the parameters of the cluster model.
- Inexact method to compute the predicted value of consumption expenditure in census data.

Implementation and Diagnostics Tests

The procedure in running the cluster model is carried out through the following steps:

1. developing the beta model (Equation 2);
2. calculating location effects (Equation 3);
3. calculating variance of estimators (Equation 4);
4. preparing the term residual to run the alpha model (Equation 6);
5. developing the generalized least squares estimate model;
6. using decomposition value singular to decompose the variance-covariance matrix as provided by the previous step to establish vectors that are randomly and normally distributed;
7. reading data census, eliminating missing values, and providing variables required by the beta and alpha models; and
8. storing all data sets required for simulation.

One of the major expected outputs of the cluster model is the headcount index (P_o), the proportion of population below a specified poverty line with reasonable reliability. Table 6.5 exhibits the summary estimation of poverty incidence for Java and non-Java provinces. As shown here, the estimation of poverty measure at provincial and district levels are reasonably reliable.

The results in Table 6.6 show that reliable poverty indicators can still be generated at the subdistrict level with standard errors of estimates less than 10 percent. At the village level, however, standard errors of estimates increased to nearly 14 percent, making them less reliable. This successful implementation was enhanced by the availability of the village census data. Complete results of the poverty mapping exercise are available from BPS Statistics Indonesia.

Finally, acceptability of the results depends on how they could be used by policy makers. However, from a technical perspective, what is desirable is a simultaneous lowering of both the level of standard errors and the level of aggregation. There is, however, a trade-off between these two goals.

Table 6.5 **Poverty Incidence (P_0) in Java and Non-Java Provinces**

Province	P_0 (%)	Interval P_0 (%), α=10%		Difference (3–4)	Standard Error
		Upper Bound	Lower Bound		
(1)	(2)	(3)	(4)	(5)	(6)
Java Provinces					
Jakarta	4.3	3.5	5.0	1.5	0.01353
West Java	19.0	18.2	19.8	1.6	0.01268
Central Java	28.4	27.8	29.1	1.4	0.01627
East Java	29.1	28.5	29.7	1.2	0.01474
Yogyakarta	26.5	25.2	27.8	2.6	0.04599
Non-Java Provinces					
Nanggroe Aceh Darussalam	13.1	11.8	14.3	2.4	0.05267
North Sumatera	17.6	16.5	18.8	2.3	0.02388
West Sumatera	11.7	10.8	12.6	1.9	0.03183
Riau	15.1	13.9	16.4	2.4	0.03325
Jambi	24.1	22.7	25.4	2.7	0.05546
South Sumatera	26.5	25.2	27.8	2.6	0.03620
Bengkulu	19.5	18.1	20.8	2.7	0.06613
Lampung	26.6	25.4	27.9	2.5	0.03475
Bangka Belitung	19.4	17.3	21.5	4.2	0.08549
Banten	12.2	11.4	12.9	1.4	0.02311
Bali	8.6	8.0	9.2	1.2	0.03142
West Nusa Tenggara	32.9	31.7	34.1	2.4	0.04728
East Nusa Tenggara	47.7	46.6	48.8	2.2	0.05610
West Kalimantan	25.4	24.4	26.4	2.0	0.04731
Central Kalimantan	16.3	15.0	17.6	2.6	0.05392
South Kalimantan	14.3	13.2	15.4	2.2	0.03955
East Kalimantan	17.7	15.7	19.7	4.0	0.04918
North Sulawesi	15.8	14.5	17.2	2.8	0.04966
Central Sulawesi	31.5	30.1	32.9	2.8	0.06812
South Sulawesi	20.3	19.4	21.1	1.7	0.03030
South East Sulawesi	32.9	31.8	34.0	2.2	0.07424
Gorontalo	23.1	20.9	25.2	4.3	0.09104

α = level of significance
Source: Authors' calculation based on poverty mapping results.

Table 6.6 **Standard Error of Poverty Incidence by Estimation Level**

	Mean Standard Error				
	Province	*District/Municipality*	*Subdistrict*	*Village*	*Total*
Java	0.00435	0.02196	0.07446	0.15967	0.14987
Non-Java	0.01019	0.02449	0.04837	0.12017	0.11380
Total	0.00900	0.02365	0.06173	0.13677	0.12921

Source: Authors' calculation based on poverty mapping results.

To test the validity of the model, Tables 6.7 and 6.8 compare P_0 as provided by the cluster estimate method and the SUSENAS, by province, in both urban and rural areas. The differences in the estimates from those provided by direct estimation which were officially published (SUSENAS) and those by census (i.e., provided by the cluster model) are almost negligible. Figure 6.5 demonstrates that the poverty estimates in rural areas produced from census data were very similar in the indices between the two approaches.

Table 6.7 **Comparison of Headcount Ratio (P_0) and Standard Error ($\hat{\delta}_{ch}$) Between Cluster Estimates and SUSENAS Results for Urban Area**

Province	*Cluster-Estimate*		*SUSENAS*		*Difference*
	P_0	$\hat{\delta}_{ch}$	P_0	$\hat{\delta}_{ch}$	(2)-(4)
(1)	(2)	(3)	(4)	(5)	(6)
Nanggroe Aceh Darussalam	10.2	0.7	10.2	3.0	0.0
Bali	7.1	0.3	9.4	2.7	(2.3)
Bangka Belitung	22.8	1.7	—	—	—
Banten	10.6	0.4	11.5	2.0	(0.9)
Bengkulu	20.5	1.2	22.0	4.5	(1.5)
Yogyakarta	21.3	0.7	23.8	3.4	(2.5)
Jakarta	4.3	0.4	4.0	0.8	0.3
Gorontalo	18.7	1.5	—	—	—
Jambi	20.0	0.9	22.4	4.3	(2.3)
West Java	19.6	0.6	18.9	2.0	0.7
Central Java	29.7	0.5	27.8	2.0	1.9
East Java	24.9	0.4	24.7	1.9	0.2
West Kalimantan	12.8	0.9	10.8	3.3	2.0
South Kalimantan	11.4	0.9	10.4	2.6	0.9
Central Kalimantan	6.8	1.3	5.6	2.5	1.1
East Kalimantan	12.8	1.4	10.0	3.9	2.9
Lampung	24.2	0.9	24.0	3.4	0.1
West Nusa Tenggara	30.4	0.9	31.9	4.2	(1.6)
East Nusa Tenggara	30.3	1.2	29.2	4.7	1.1
Riau	9.0	0.7	9.1	2.8	(0.0)
South Sulawesi	15.4	0.4	18.3	3.3	(2.9)
Central Sulawesi	21.2	0.8	23.1	6.0	(1.9)
South East Sulawesi	15.0	0.5	15.7	5.6	(0.7)
North Sulawesi	11.2	1.2	—	—	—
West Sumatera	17.4	0.8	18.2	3.9	(0.8)
South Sumatera	24.2	1.2	—	—	—
North Sumatera	18.0	0.9	18.3	2.5	(0.3)

SUSENAS = National Socioeconomic Survey
Source: Authors' calculation based on Poverty mapping results.

To ensure the validity and reliability of the models, a diagnostic test was done as illustrated in Table 6.9. The table shows the results for Nanggroe Aceh Darussalam–Urban, on which there are two major points worth mentioning. First, the model is able to explain some 50 percent variation of headcount index, that is, 0.50. Second, the multiplication of the mean and model parameter (i.e., the regression coefficient) for each variable is very similar between the two sources, for both unweighted and weighted versions. For an inspection, it is useful to focus on the sums of the products between the two sources. The sum for the weighted version, for example, is 11.94[6] and for poverty mapping (according to the population census or Sensus Penduduk–SP 2000) it is 11.95 (equivalent to Rp154,817[7]).

For further inspection, a visual presentation of the distributions of the consumption models derived from the SUSENAS and the census is provided

6 This is about Rp153,277; equal to the average value of logarithmic per capita expenditure, according to the SUSENAS.

7 Rp stands for rupiah.

Table 6.8 **Comparison of Headcount Ratio (P_0) and Standard Error ($\hat{\delta}_{ch}$) Between Cluster Estimates and SUSENAS Results for Rural Area**

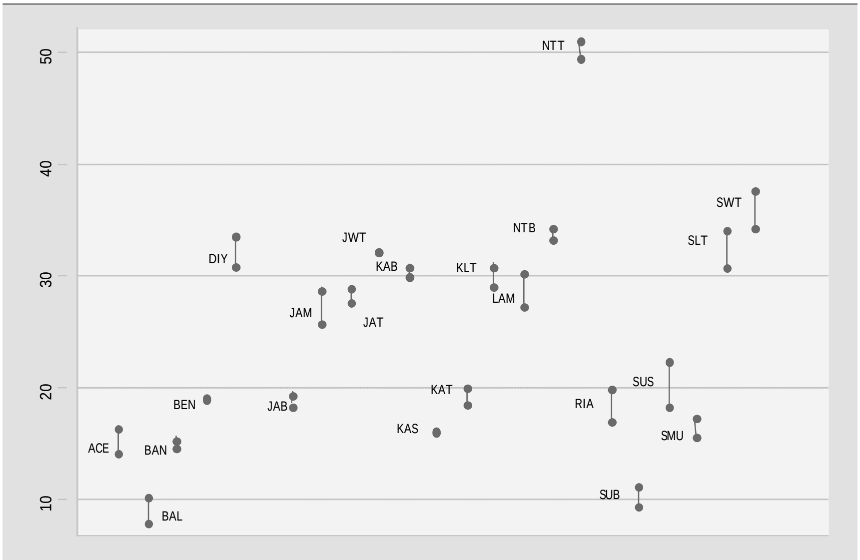
Province	*Cluster-Estimate*		*SUSENAS*		*Difference*
	P_0	$\hat{\delta}_{ch}$	P_0	$\hat{\delta}_{ch}$	(2–4)
(1)	(2)	(3)	(4)	(5)	(6)
Nanggroe Aceh Darussalam	14.2	0.8	16.3	2.8	(2.1)
Bali	10.2	0.4	7.9	1.8	2.3
Bangka Belitung	16.9	1.0	...	...	...
Banten	14.6	0.6	15.4	1.4	(0.8)
Bengkulu	19.0	0.7	18.9	4.8	0.2
Yogyakarta	33.6	0.9	30.8	3.3	2.8
Jakarta	...	...	...	...	...
Gorontalo	24.6	1.2	...	...	...
Jambi	25.7	0.7	28.6	5.4	(2.9)
West Java	18.4	0.4	19.3	1.4	(0.9)
Central Java	27.6	0.4	28.8	1.6	(1.2)
East Java	32.0	0.3	32.1	1.6	(0.1)
West Kalimantan	29.9	0.6	30.7	3.3	(0.8)
South Kalimantan	16.0	0.6	16.2	2.7	(0.2)
Central Kalimantan	20.0	0.6	18.5	4.2	1.4
East Kalimantan	29.0	1.2	30.7	4.9	(1.7)
Lampung	27.3	0.7	30.2	3.3	(3.0)
West Nusa Tenggara	34.2	0.7	33.2	3.0	1.0
East Nusa Tenggara	50.9	0.6	49.4	3.7	1.5
Riau	19.8	0.8	17.0	3.4	2.9
South Sulawesi	22.3	0.6	18.4	2.5	4.0
Central Sulawesi	34.0	0.9	30.7	4.6	3.4
South East Sulawesi	37.6	0.6	34.2	5.6	3.4
North Sulawesi	18.5	0.6	...	...	...
West Sumatera	9.4	0.5	11.2	2.0	(1.9)
South Sumatera	27.7	0.6	...	...	...
North Sumatera	17.3	0.5	15.5	2.2	1.8

SUSENAS = National Socioeconomic Survey
Source: Authors' calculation based on Poverty mapping results.

(Figures 6.6 and 6.7). These figures provide a visual presentation of the results by comparing the distributions of estimates from SP 2000 with SUSENAS 1999. Results for the province Nanggroe Aceh Darussalam, urban and rural areas, are used as examples.

The comparisons show that expenditure from the SUSENAS is slightly lower than expenditure from SP 2000 in both urban and rural areas. For urban areas, the distributions fit each other within the interval of 6–50 cumulative percent, but then SP 2000 produced higher results within the interval of 50–90 percent. Beyond that, SUSENAS produced higher percentage results. For rural areas, the distributions are the same within the interval of 6–40 cumulative percentages and higher for SP 2000 for the rest of the percentages. Overall, the distributions of the two results for all provinces under study fit each other relatively well. As far as the headcount index is concerned, the most important is the distribution of the results for the lowest 30 percent of the income distribution as the headcount ratio is within this range.

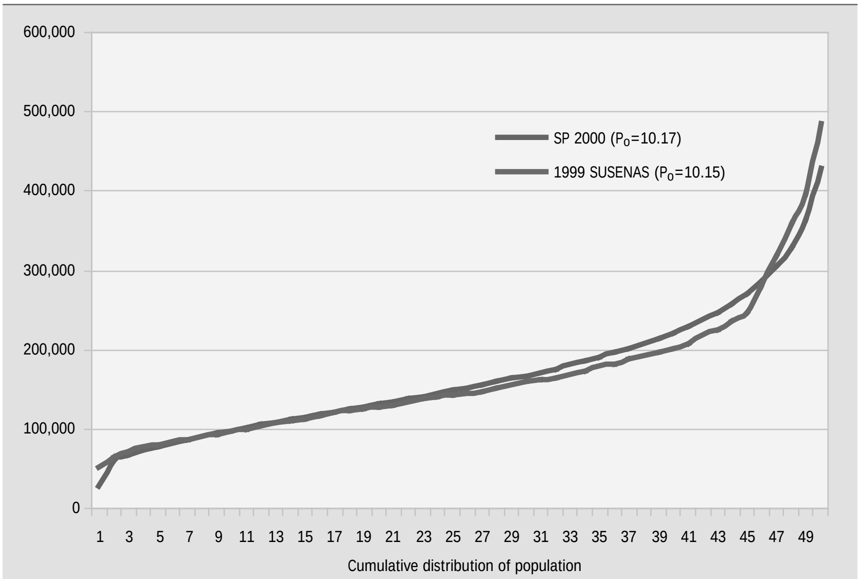
Figure 6.5 **Comparisons of Poverty Estimates Between the Cluster-Method and the SUSENAS in Rural Areas, 2000**



ACE = Nanggroe Aceh Darussalam; BAL = Bali; BAN = Banten; BEN = Bengkulu; DIY = D. I. Yogyakarta; JAB = Jawa Barat; KAS = Kalimantan Selatan; KLT = Kalimantan Timur; KAT = Kalimantan Tengah; LAM = Lampung; NTB = Nusa Tenggara Barat; NTT = Nusa Tenggara Timur; RIA = Riau; SUB = Sumatera Barat; SMU = Sumatera Utara; SUS = Sulawesi Selatan; SLT = Sulawesi Tengah; SWT = Sulawesi Tenggara

Source: Authors' calculation based on Poverty Mapping Results.

Figure 6.6 **Percentage Distribution of Expenditure in Nanggroe Aceh Darussalam—Urban Area**



SP = Population census.

Source: Authors' calculation based on poverty mapping and SUSENAS results.

Table 6.9 **Diagnostic Tests of Nanggroe Aceh Darussalam—Urban Area**

Nanggroe Aceh Darussalam–Urban

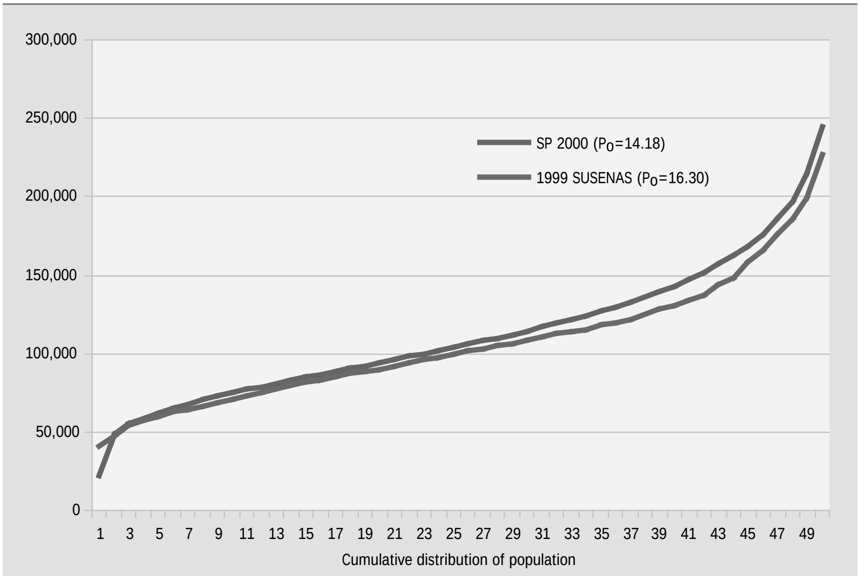
	Unweighted Mean		*Weighted Mean*		*Parameter*	*Unweighted Mean x (b)*		*Weighted Mean x (b)*	
Variable Name	*SUSENAS 1999*	*SP 2000*	*SUSENAS 1999*	*SP 2000*	*(b)*	*SUSENAS 1999 (2)x(6)*	*SP 2000 (3)x(6)*	*SUSENAS 1999 (4)x(6)*	*SP 2000 (5)x(6)*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
thhsize	4.46	4.12	5.59	5.04	-0.23233	-1.04	-0.96	-1.30	-1.17
vsecth3	0.32	0.36	0.34	0.36	1.12880	0.37	0.41	0.38	0.41
vwork	0.81	0.85	0.85	0.85	0.49844	0.41	0.42	0.42	0.42
hhs_prad	1.47	1.42	1.78	1.60	0.10169	0.15	0.14	0.18	0.16
vcba	2.75	2.93	2.89	2.99	-0.23723	-0.65	-0.70	-0.69	-0.71
veduch4	0.14	0.12	0.13	0.12	1.55913	0.22	0.19	0.21	0.19
sex	0.81	0.86	0.87	0.89	0.12695	0.10	0.11	0.11	0.11
thhsize2	24.89	20.75	35.69	29.31	0.01142	0.28	0.24	0.41	0.33
constant					12.20751	12.20751	12.20751	12.20751	12.20751
					R-squared=50.0%	12.05	12.06	11.94	11.95

Nanggroe Aceh Darussalam–Rural

	Unweighted Mean		*Weighted Mean*		*Parameter*	*Unweighted Mean x (b)*		*Weighted Mean x b()*	
Variable Name	*SUSENAS 1999*	*SP 2000*	*SUSENAS 1999*	*SP 2000*	*(b)*	*SUSENAS 1999 (2)x(6)*	*SP 2000 (3)x(6)*	*SUSENAS 1999 (4)x(6)*	*SP 2000 (5)x(6)*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
rasio	0.87	0.84	0.91	0.90	-0.12957	-0.11	-0.11	-0.12	-0.12
hhsize	4.54	4.25	5.45	5.10	-0.07833	-0.36	-0.33	-0.43	-0.40
married	0.81	0.86	0.87	0.91	0.08726	0.07	0.08	0.08	0.08
ussch	0.06	0.08	0.06	0.08	-0.17671	-0.01	-0.01	-0.01	-0.01
health	0.28	0.35	0.28	0.35	0.71542	0.20	0.25	0.20	0.25
dist_ls	0.53	0.46	0.52	0.46	0.06087	0.03	0.03	0.03	0.03
elsch	0.63	0.73	0.64	0.73	-0.15090	-0.10	-0.11	-0.10	-0.11
comm	0.15	0.18	0.15	0.19	-0.55923	-0.08	-0.10	-0.08	-0.11
age_rasio	39.45	35.96	41.01	38.60	0.00196	0.08	0.07	0.08	0.08
vsex	0.87	0.88	0.87	0.88	-4.76834	-4.15	-4.21	-4.15	-4.21
vage	43.53	42.33	43.68	42.39	-0.01692	-0.74	-0.72	-0.74	-0.72
vhhsize	4.16	4.25	4.18	4.32	0.91611	3.81	3.90	3.83	3.96
vmarried	0.85	0.86	0.85	0.86	3.02091	2.57	2.60	2.57	2.60
veduch1	0.68	0.65	0.68	0.65	-11.57397	-7.86	-7.52	-7.90	-7.54
veduch2	0.16	0.17	0.16	0.17	-12.49233	-2.00	-2.09	-2.01	-2.08
veduch3	0.13	0.16	0.13	0.16	-9.92067	-1.30	-1.57	-1.27	-1.56
tssch	0.09	0.01	0.09	0.01	0.72027	0.06	0.01	0.06	0.01
vsecth2	0.09	0.09	0.09	0.09	-0.94309	-0.08	-0.08	-0.08	-0.08
vsecth3	0.11	0.10	0.11	0.10	-2.38987	-0.26	-0.23	-0.26	-0.23
vwkstath1	0.42	0.48	0.43	0.48	1.47497	0.63	0.71	0.64	0.71
vwkstath2	0.31	0.28	0.31	0.28	1.60297	0.50	0.44	0.49	0.45
vwkstath3	0.17	0.15	0.17	0.15	1.91081	0.33	0.29	0.32	0.29
vcba	3.20	3.25	3.22	3.31	-0.09352	-0.30	-0.30	-0.30	-0.31
pr_telp	0.01	0.01	0.01	0.01	-18.94475	-0.19	-0.13	-0.20	-0.13
vrasio	0.85	0.84	0.85	0.85	-2.24346	-1.90	-1.89	-1.90	-1.90
vprsckid	0.19	0.20	0.19	0.20	-4.77307	-0.90	-0.96	-0.90	-0.97
vprunde5	0.09	0.10	0.09	0.10	-3.84777	-0.36	-0.40	-0.36	-0.40
vownhou	0.61	0.62	0.62	0.62	0.21653	0.13	0.13	0.13	0.13
vrenthou	0.03	0.02	0.03	0.02	0.64336	0.02	0.01	0.02	0.02
distkec	7.00	11.02	6.96	10.91	-0.01951	-0.14	-0.21	-0.14	-0.21
density	1.97	2.35	1.97	2.34	0.09461	0.19	0.22	0.19	0.22
skor	5.08	4.60	5.09	4.61	0.03337	0.17	0.15	0.17	0.15
vilsect1	0.98	0.99	0.98	0.98	-2.05431	-2.01	-2.02	-2.00	-2.02
constant					25.65781	25.65781	25.65781	25.65781	25.65781
					R-squared=61.0%	11.60	11.55	11.53	11.52

SUSENAS = National Socioeconomic Survey; SP = Census of population
Source: Authors' calculation based on the poverty mapping results.

Figure 6.7 **Percentage Distribution of Expenditure in Nanggroe Aceh Darussalam—Rural Area**



SP = Population census.
Source: Authors' calculation based on poverty mapping and SUSENAS results.

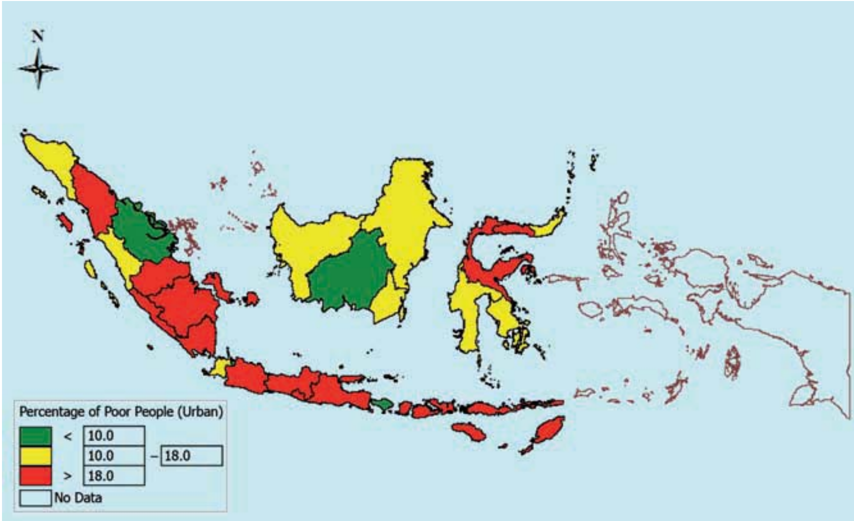
Developing a GIS Application of the Results

Recent studies on cluster estimation overlaid by thematic maps offer a promising avenue for analyzing the potential poverty impact of a variety of policy proposals. One could look into, for example, the potential impact of geographically targeted transfer schemes (Yin et al. 2004). All cluster-model results discussed in this chapter have been presented in thematic maps such as the map in Figure 6.8. They are generated through a dynamic, flexible, and user-friendly type of GIS application named PRISMA, or Poverty Reduction Information System for Monitoring and Analysis. A complete description of PRISMA, including examples of its application, is presented in the appendix of this chapter.[8]

PRISMA interactively combines district poverty indicators at household and population levels with other poverty-related indicators such as population density, share of agriculture by household, communication facilities, access to TV, access to school (secondary and high school), access to hospital, access to electricity, access to a safe-water facility, average urban score, welfare status, and average distance to the center of the subdistrict.

8 A CD-ROM version of PRISMA can be obtained from ADB's Economics and Research Department.

Figure 6.8 **Percentage of Poor Population in Urban Areas by Province**



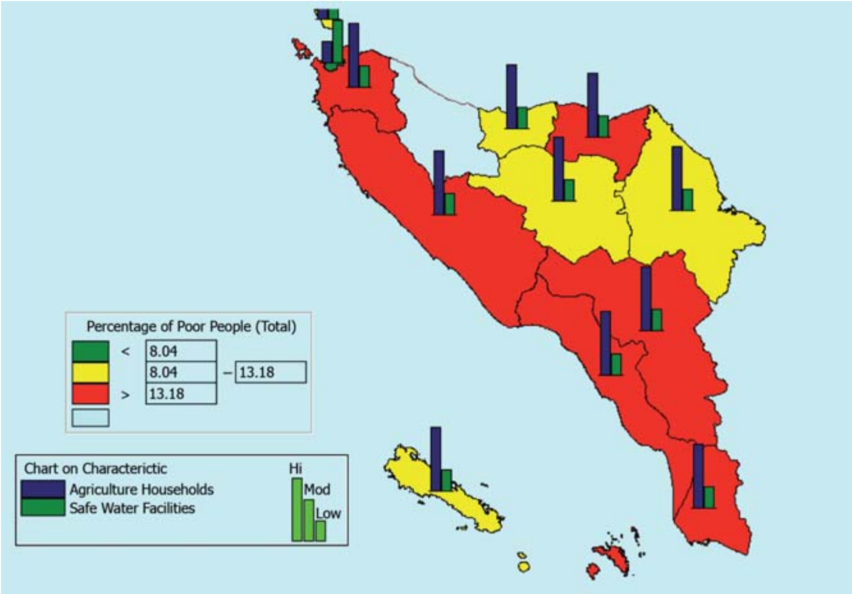
Note: The map that presents the geographical distribution of poor and nonpoor based on the poverty mapping results.
Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

In the system, the poverty indicators maps are presented using the traffic-light classification system (see Figure 4.8) mentioned earlier, in which red represents high, yellow average to moderate, and green low poverty incidence. The absence of color in an area on the map indicates that data is not available for that particular area. Geographic targeting can thus be visually illustrated according to the information available. This figure shows, for example, that the lower part of Indonesia (from North Sumatera to East Nusa Tenggara) is comparatively poorer based on the poverty headcount criterion of above 18 percent.

In addition to the default cutoff points that represent actual results from poverty mapping, users can also change the cutoff points and do spatial analysis using other levels of poverty incidence. Other features include the overlaying of bar charts of poverty characteristics, altering the traffic-light classification, presenting detailed information about a province or district, exporting maps for use in other software applications, and printing output.

Figure 4.9 is an example of how some socioeconomic variables can be overlaid on the poverty map. In addition to indicating poverty incidence in Nanggroe Aceh Darussalam using the traffic-light classification system, data from the Podes on the proportion of agricultural households and access to safe-water facilities is overlaid on the poverty map to show that a high proportion of households in the province are agricultural while access to safe-water facilities is moderate in all districts except in Banda Aceh and Sabang districts.

Figure 6.9 **Percentage of Poor People in Nanggroe Aceh Darussalam Province with Some Overlaying Variables by District**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

Overlying Variables

This section discusses variables used to overlay poverty incidence based on the headcount index in the GIS application. These "layering" variables of the poverty mapping result are correlates of poverty identified in the 2000 Podes survey. The unit of observation is the village, which is aggregated into district and municipality levels to be consistent with the district level measured for the headcount ratio.

To emphasize the user-friendly characteristics of the system, the cutoff points of the variable can be changed by the users according to their interests or concerns. For example, the default criteria for good access to hospital facilities is: 75 percent or more of the villages in a district must have their own hospital or are located not farther than 2.5 kilometers from the hospital. The system allows users to change this threshold, i.e., from 75 percent to 50 percent, for instance.

The thresholds used to categorize variables are set up differently across provinces, such as the distance from the village to the subdistrict capital. The reason for this is that population density and distribution vary considerably across provinces. Five kilometers to the subdistrict capital is considered relatively far in Java but not so in other provinces outside Java. Table 6.10 lists the variables with their different thresholds. For example, in the case of North Sumatera, less than three kilometers from the subdistrict capital city is considered close, while more than seven kilometers is considered far. The rationale for this is that 25 percent of villages in North Sumatera are located less than three kilometers from the capital city of their respective subdistricts, while 50 percent of them are located between three and seven kilometers, and 25 percent are more than seven kilometers. The capital city of a subdistrict is used as a reference because some basic public facilities like the public health center (*Puskesmas*) and junior and senior high schools are usually located in the capital city of a subdistrict.

Table 6.10 **Thresholds Used for Classifying Distances from Village to Subdistrict Capital by Province** (in kilometer)

Province	*Close* [a]	*Far* [b]
Nanggroe Aceh Darussalam	2.0	4.0
North Sumatera	3.0	7.0
West Sumatera	2.0	3.8
Riau	3.5	10.0
Jambi	3.0	9.0
South Sumatera	4.0	10.0
Bengkulu	3.0	6.0
Lampung	3.0	6.0
Bangka Belitung	2.9	9.0
Jakarta	1.2	2.0
West Java	2.0	4.0
Central Java	2.0	4.0
Yogyakarta	1.5	3.0
East Java	2.0	4.0
Banten	2.4	5.0
Bali	2.5	5.0
West Nusa Tenggara	1.5	4.0
East Nusa Tenggara	4.2	10.0
West Kalimantan	5.0	13.0
Central Kalimantan	7.0	20.0
South Kalimantan	2.5	5.0
East Kalimantan	4.1	14.5
North Sulawesi	1.9	5.0
Central Sulawesi	4.0	12.0
South Sulawesi	2.5	6.0
South East Sulawesi	3.0	8.0
Gorontalo	2.0	4.0

a = The lowest quintile (the closest 25%)
b = The highest quintile (the farthest 25%)
Source: Authors' calculation.

The sensitivity of the proposed layer variables is examined by observing variation in the headcount index between categories. For example, the percentage of agricultural households (Agric) is correlated with the headcount ratio, the overlying index is found to vary with the Agric variable by 14 percent in the lowest category, 21 percent in the medium category, and 26 percent in the highest category. In other words, the proportion of agricultural households, to some extent, explains variation in the headcount index–the higher the proportion, the higher the index. Tables 6.11 and 6.12 highlight the test results of the sensitivity of the variables concerned.

Conclusion

Poverty indicators derived from household surveys on income or consumption, or both, have a limited regional disaggregation. In this study, poverty mapping modeling is implemented by using household surveys and population census to estimate poverty indicators down to the smallest administrative units, i.e., for district to village levels. The methods have been

Table 6.11 **Categorization of Layer Variables in the GIS Application of Poverty Mapping Results**

Variable Name	*Label*	*Indicator*	*Category*	*Number of Districts*	*Average P_0 (%)*	*Std. Dev. of P_0*
Urban	Urban score	Composite index of urban	Low Urban	87	0.276	0.116
			Urban	89	0.219	0.089
			High urban	142	0.199	0.099
Density	Population density	Population per square kilometer	Low	98	0.252	0.108
			Medium	114	0.237	0.100
			High	106	0.189	0.101
Agric	Agriculture households	Percentage of agriculture households	Low	50	0.135	0.075
			Medium	94	0.208	0.082
			High	174	0.261	0.108
TelCom	Communication facilities	Percentage of villages with communication facilities	Low	85	0.273	0.114
			Medium	124	0.234	0.094
			High	109	0.180	0.093
TV	TV	Percentage households having TVs	Low	83	0.304	0.117
			Medium	207	0.210	0.081
			High	28	0.110	0.073
ScSch	Access to secondary school	Percentage of villages having secondary school or located 2.5 km or less	Low	3	0.306	0.148
			Medium	224	0.249	0.102
			High	91	0.166	0.091
HgSch	Access to high school	Percentage of villages having high schools or located 2.5 km or less	Low	102	0.272	0.114
			Medium	158	0.226	0.091
			High	58	0.145	0.076
Hospital	Access to hospital	Percentage of villages having hospitals or located 2.5 km or less	Low	251	0.249	0.102
			Medium	57	0.143	0.072
			High	10	0.127	0.071
Poor	Poor family	Percentage households considered as under welfare	High	45	0.119	0.058
			Medium	261	0.234	0.092
			Low	12	0.441	0.114
Electr	Electricity	Percentage of households using electricity	Low	13	0.420	0.114
			Medium	196	0.247	0.093
			High	109	0.164	0.083
Water	Safe water facilities	Percentage households using pipe or pump-water facilities	Low	222	0.254	0.100
			Medium	69	0.175	0.093
			High	27	0.124	0.071
Distance	Distance to center of subdistrict	Percentage of villages by distance to center of subdistrict office	Low	60	0.154	0.0687
			Medium	42	0.201	0.1152
			High	216	0.251	0.1028

Std. Dev. = Standard Deviation
Notes: The first and the highest quintiles are used for the categorization except otherwise stated and P_0 as head count index in percent.
Source: Authors' calculation based on the poverty mapping results.

Table 6.12 **Pearson Correlations among Layered Variables and between Layered Variables and Headcount Ratio (P_0)**

	DENSITY	*AGRIC*	*TELCOM*	*TV*	*SCSCH*	*HGSCH*	*HOSPIT*	*URBAN*	*POOR*	*ELECTR*	*WATER*	*DISTANC*
P0	-0.36	0.49	-0.37	-0.62	-0.36	-0.44	-0.42	-0.45	0.73	-0.58	-0.45	0.37
DENSITY		-0.82	0.61	0.61	0.65	0.76	0.81	0.87	-0.37	0.55	0.72	-0.56
AGRIC			-0.79	-0.81	-0.81	-0.91	-0.90	-0.97	0.50	-0.73	-0.76	0.71
TELCOM				0.78	0.86	0.83	0.68	0.84	-0.44	0.81	0.65	-0.51
TV					0.75	0.78	0.72	0.82	-0.64	0.87	0.67	-0.55
SCSCH						0.93	0.76	0.87	-0.41	0.75	0.63	-0.60
HGSCH							0.89	0.94	-0.47	0.74	0.73	-0.71
HOSPIT								0.91	-0.43	0.64	0.76	-0.73
URBAN									-0.48	0.77	0.78	-0.69
POOR										-0.59	-0.43	0.35
ELECTR											0.64	-0.49
WATER												-0.57

Note: All bivariate correlations are significant at one per cent level (2-tailed).
Source: Authors' calculation based on the poverty mapping results.

implemented successfully in a number of countries. The technique can also be used to generate other welfare indicators such as the welfare index, nutrition status, basic needs index, school drop-out rate, and inequality measures.

The application of poverty mapping in Indonesia incorporates information from the Podes to strengthen the modeling results. The overall results show that the poverty mapping technique can generate reliable poverty indicators at district and subdistrict levels with standard errors estimates of less than 10 percent. In some cases, the estimation can actually go down to the village level, but the estimates at the village level are generally less reliable as their standard errors reach about 14 percent. The successful implementation of poverty mapping brings with it a reminder to make more use of the census data, which seems still underutilized in most developing countries. Poverty mapping results of this study were also used as a basis for a GIS application by combining with other poverty-related information in a dynamic interactive PRISMA.

Appendix 6.1

Poverty Reduction Information System for Monitoring and Analysis: A GIS Application of Poverty Mapping Results

Guntur Sugiyarto, Dudy Sulaeman, Eric B. Suan, and Mary Ann Magtulis.

Introduction

Estimation of poverty indicators at a more disaggregated geographical area is implemented in Indonesia by using a poverty mapping technique. The estimation is conducted by using data sets from three sources, namely, the household expenditure survey (SUSENAS), village census (Podes), and population census (Sensus Penduduk–SP) data. The technique maximizes the rich information of surveys and the wider coverage area of censuses. The results basically show that the poverty indicator estimates are reliable even at the village level in Java; while for outside Java, the estimates are only reliable up to the subdistrict level.

However, statistical tables may not be as revealing and intelligible to most people as they should be–not even to regular data users. Thus, a geographic information system (GIS) application was developed by incorporating poverty indicator estimates for small areas such as districts with other poverty-related information. The geographically disaggregated poverty indicators are used to provide information on the spatial distribution of poverty. This information can be used as a decision-support system for specific evidence-based interventions, programs, and plans for targeting the poor (Albert et al. 2003).

This report summarizes the development of a GIS tool that could display geographically referenced information (i.e., spatial data) of poverty characteristics and create visuals of meaningful relationships and significant patterns in data. The tool is called the Poverty Reduction Information System for Monitoring and Analysis (PRISMA).

PRISMA allows users to simulate changes in poverty incidences to reflect different level of targets that are regularly faced by developing countries like Indonesia. It can therefore provide meaningful information for monitoring and analysis. The system adopts a "traffic-light" classification system of red, yellow, and green to represent, respectively, high, average, and low poverty incidences.

The construction of interactive poverty-referenced maps helps in visualizing disparities of living standards across regions. This visual information is useful in identifying areas that need additional resources for poverty reduction. A causal relationship between the welfare status of households and geographic or other factors may be displayed. As a result, improved poverty targeting may be better planned. The provinces, districts, subdistricts, and even villages where the poor households are located, for instance, may be selected for some programs such as to improve infrastructure and education and health facilities. These areas may also be targeted for direct transfer programs such as food-for-work, improved access to credit, or direct government subsidies to enhance the availability of social services to those who need them most.

Poverty Reduction Information System for Monitoring and Analysis

PRISMA was developed by using two computer software programs– MapObject 2.1 and Visual Basic 6.0. The system runs on Windows XP Professional. It has a comprehensive database of spatial information based on the poverty mapping results and other sources. For the Indonesian data set, however, spatial information provided by PRISMA is available for only 27 out of 30 provinces of Indonesia. This is because SUSENAS 1999, one of the sources of data sets used in the small-area estimation of Indonesia's poverty indicators, covered only these 27 provinces. Excluded provinces are Maluku, Maluku Utara, and Irian Jaya, which is now known as Papua.

The system is user-friendly and very intuitive as it is very easy to run and understand. It has standard geographic data and other spatial information to ensure universal compatibility and replicability for other countries. The tool was pilot-tested by using poverty mapping modeling results conducted in Indonesia that can be scaled for other countries.

Users can view thematic maps showing spatial distribution of one or more specific data themes for a particular geographic area. Data themes that can be generated using PRISMA menus are: spatial disaggregation, and population, household, and poverty characteristics related to Indonesia. Other PRISMA features include the overlay of bar charts of poverty characteristics, flexible alteration of the traffic-light classification of thematic maps, presentation of detailed information about a province or district, export of maps for use in other software application, and output printing.

How to use PRISMA

Figure 1 shows PRISMA's opening screen: the provincial map of Indonesia with an embedded overview map. The top of the screen has a drop-down menu for map disaggregation with submenus on *population*, *household,* and *characteristics*. Other features include GIS functions that allow users to view more detailed information about the selected area, zoom in and out, move the map around to review its perimeter (when zoomed in), revert to the original map size, and print.

Appendix Figure 6.1 **Introductory Screen of PRISMA**

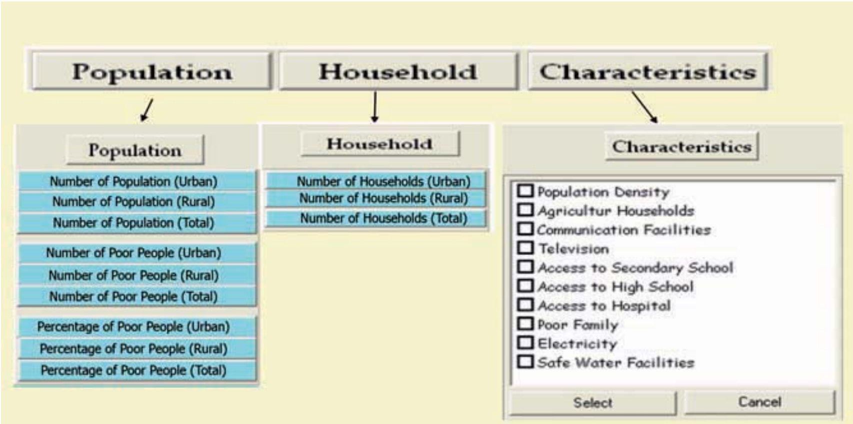
Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

Viewing Poverty Maps

To view spatial information in a map, users choose the level of administrative aggregation–national to district levels–from the drop-down menu. Specifically, users can choose a map of Indonesia with provincial or district data, and a map of a selected province with disaggregated information on districts.

To view poverty indicators of a province or district, choices are listed on the population and household menus, which can then be combined with indicators available on the characteristics menu. Appendix Figure 6.2 shows the detailed indicators available in each menu.

Appendix Figure 6.2 **Menu Bars for Population, Households, and Characteristics**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

The population menu contains spatial information on the sizes of populations and the number and percentage of poor people in rural and urban areas, and in total. The household menu shows the number of households in urban and rural areas, and in total. The characteristics menu provides information on the following indicators:

- Population density: number of people per square kilometer
- Agriculture households: percentage of households whose head's primary occupation is in agriculture
- Communication facilities: percentage of villages with communication facilities such as telephone and fax lines
- TVs: percentage of households with TV sets
- Access to secondary schools: percentage of villages with a secondary school located within its vicinity or at a radius of not more than 2.5 kilometers (km)
- Access to high schools: percentage of villages with a high school located within its vicinity or at a radius of not more than 2.5 km
- Access to hospitals: percentage of villages with a hospital located within its vicinity or at a radius of not more than 2.5 km
- Urban score: total score of the composite urban index for the village–the higher the value, the more urban the area
- Under-welfare family: percentage of households considered under-welfare based on the welfare classification developed by the National Coordinating Board for Family Planning
- Electricity: percentage of households with access to electricity

- Safe-water facilities: percentage of households with access to a water pipe or pump
- Distance to the center of the subdistrict: percentage of villages by distance to the center of the subdistrict office (subdistrict capital)

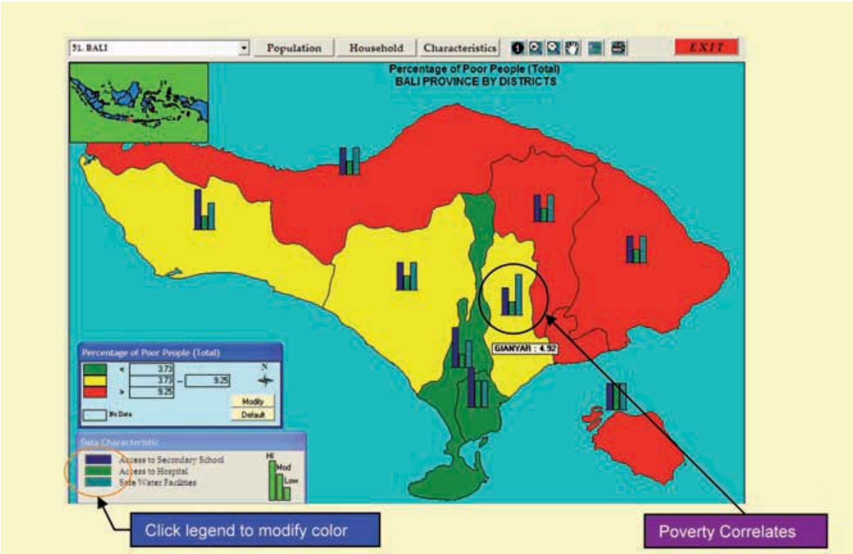
The characteristics menu cannot be activated, however, if the map chosen at the drop-down menu is *Indonesia by Provinces* or *Indonesia by Districts.* The indicators are not visible at these levels and cannot be visually presented in those maps. The combination of poverty indicators at the household or per capita level with other indicators available in the characteristics menu (described below) can only work on maps of individual provinces disaggregated by districts.

The population or household menu contains a poverty indicator theme presented in a three-colored map–using the traffic-light classification system of poverty indicators. Green areas connote the lowest magnitude or below-average poverty regions, yellow portrays regions with moderate or average poverty, and red represents the highest magnitude or above-average poverty regions. Regions with no color on the map indicate that there is no data available for that particular area.

The poverty indicator theme map can then be combined or overlaid with one or more other indicators available in the characteristics menu. This overlying system can be used to examine the association of poverty indicators with other indicators. These indicators will overlay the poverty indicator map theme with bar charts which indicate high, moderate, and low scales–as defined in a legend–of the selected indicators. Users can change the color, move, and even resize the legends to improve the presentation.

These features thus allow geographic targeting to be visually illustrated according to the information provided by the poverty mapping results, which can be enhanced by overlaying other indicators from other sources such as the Podes. Appendix Figure 6.3, for example, shows the percentage of poor people in urban and rural districts of Bali province using the traffic-light classification scheme of the poverty indicators as the spatial theme. Bar charts of access to secondary schools, hospitals, and safe-water facilities are overlaid on the district map. The result shows that poverty incidence seems to be concentrated in the northern part of the island. Access to safe-water facilities is relatively good and in one district, i.e., Gianyar, the access rate to safe-water facilities is even better than access to education.

Appendix Figure 6.3 **Poverty Indicators Based on the Traffic Light Classification System Overlaid with Bar Charts of Other Important Variables**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

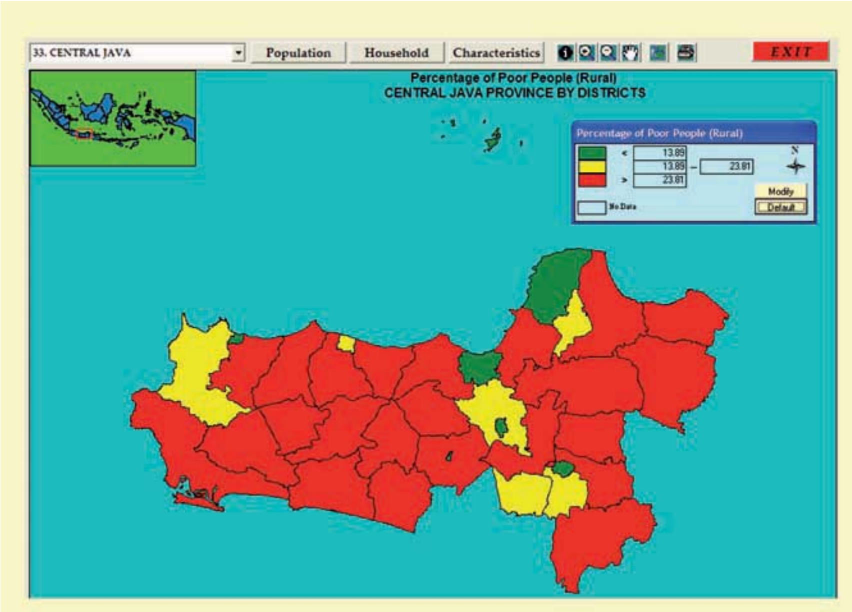
Modifying the Classification

The "default" settings for each specific subject in PRISMA are arbitrary, making PRISMA flexible and user-friendly. Aside from viewing the map, the user can also modify the default classification of the poverty condition by changing the legend of the traffic-light classification system. The user can alter the value in the interval of classification and click on *modify* to activate the change. The new cutoff points display a different level of grouping and automatically change the color distribution of the map. Clicking on *default* reverts the image to one showing the default upper or lower limit of the interval. Appendix Figure 6.4 and 6.5, for example, show the percentage of poor people in rural areas in Central Java. Appendix Figure 6.4 follows the default traffic-light color distribution, while Appendix Figure 6.5 displays a different color distribution after the yellow interval's upper limit was changed from 23.81 to 25 percent. This change increased the number of districts in yellow and diminished those in red.

Using the Information Icon

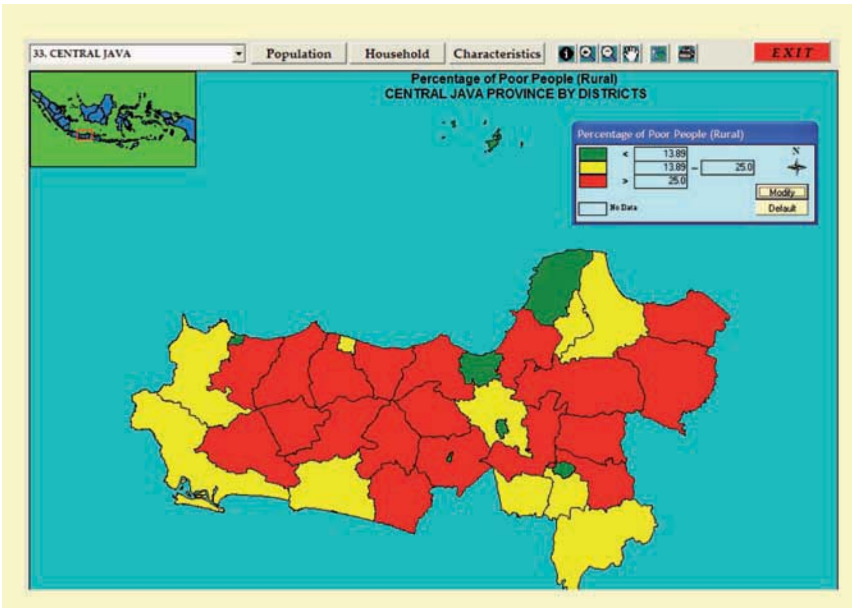
The information icon, , provides poverty details of an area. By pointing the cursor to the interactive map and clicking on an area of interest, a new window is displayed showing a statistical table and charts. The table presents

Appendix Figure 6.4 **Default Classification of the Poverty Incidence**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

Appendix Figure 6.5 **Modified Classifications of the Poverty Incidence**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

values of variables chosen from the population, household, and characteristics menus. These are the same variables on the menus of the introductory screen (the characteristics menu is not activated for the provincial level). The bar chart below the table shows the graphical distribution of the districts. Its theme depends on the variable chosen from the table above it, and the theme is implemented by clicking on the variable name.

The user can also create a graph of the variable of interest by clicking on the checkbox left of the variable name. The resulting graph appears on the table's right. The user can click on more than one variable to compare poverty statistics of the district or province under review.

There is also an option to either print the window in view or to go back to the main menu. The print option copies the table or graphs to a digital "clipboard" for pasting in other software applications as a picture object. In this way they can be printed on paper. (See Printing the Map below.)

As shown in Appendix Figure 6.6, by clicking on the Musi Banyu Asin district (where 27.22 percent of the total population is poor) in the map of South Sumatera (or Sumatera Selatan) province, a new window appears. The statistical table in the upper left of the new window shows the poverty characteristics of the district. The bar chart on the table's right shows that a low percentage of villages in Musi Banyu Asin have communication facilities but that a moderate percentage of households have TV sets. The chart in the lower portion shows that the Musi Banyu Asin district is only second among districts in Sumatera Selatan when it comes to under-welfare families, the highest is found in Ogar Komering Ilir, and the lowest is in Muara Enim.

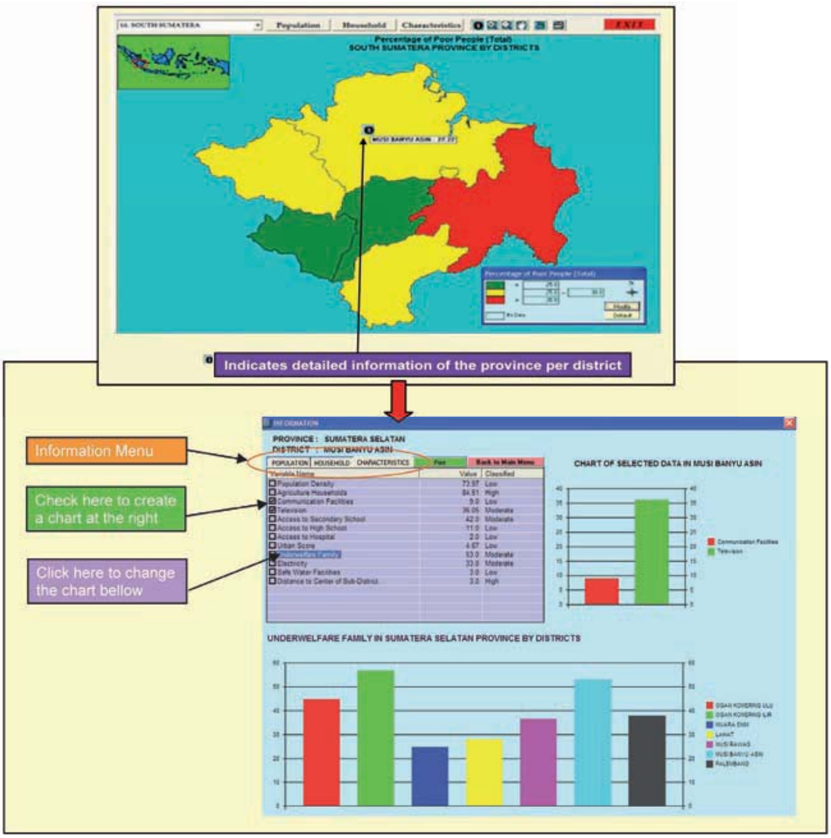
Other GIS Icons

Zooming In, *Zooming Out*, *Full Extent,* and *Pan Map* tools are used to change the magnification of the map. When the mouse is dragged to any side of the window, magnification increases (zooming in). Clicking any space on the map triggers zooming out. The Full Extent tool reverts the map to its original size. The Pan or Hand Map tool is used to move the map around to view its perimeter and is used only if the map is already zoomed in. Appendix Figure 6.7 shows, for example, by zooming in on a map of Southeast Sulawesi, the number of poor people in the rural areas of the province's Kendari and Muna districts is displayed.

Printing the Map

The print bar allows the user to change the layout of the map and use it in other computer applications. Appendix Figure 6.8 displays the map of Jakarta

Appendix Figure 6.6 **Displaying the Related Statistical Tables and Graphs Using the Information Window**



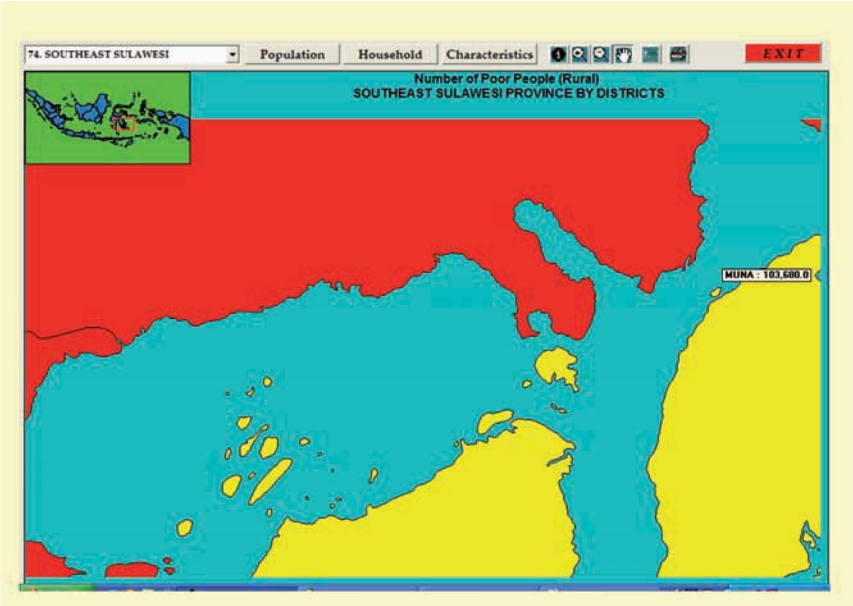
Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

province by districts in the print menu environment, indicating the number of households in urban areas. Here, the user can alter the default layout by changing the background color of the map, presence of the north-orientation graphic, traffic-light classification, and legend of chart or data characteristics. The user can also move the position of the map title and other parts of the map. When the layout is final, the user can view the output by clicking the *Preview* button.

The *Hint* button reveals guidelines or tips on how to correctly print the map. The following are statements found on this dialog box:

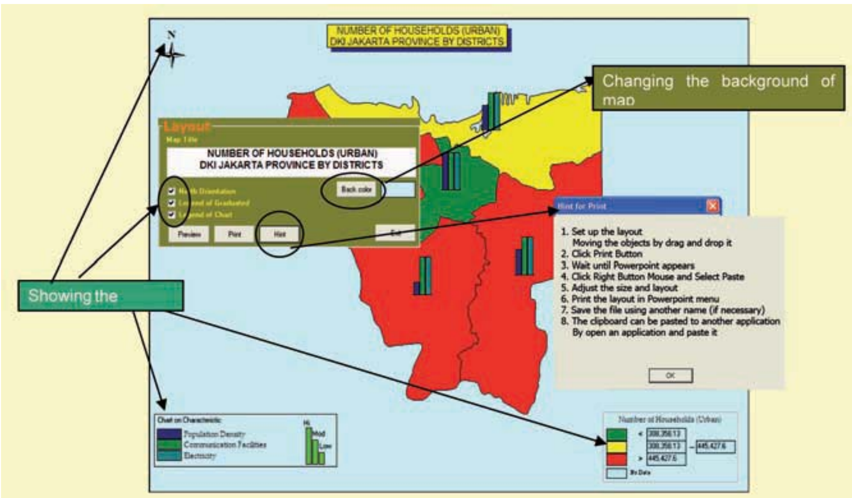
- Set up the layout. Move objects by dragging and dropping them–this changes the general appearance of the map.
- Click print button. This does not print out the map, rather, the map is copied onto the clipboard.

Appendix Figure 6.7 **Example of Zooming in a Map of Southeast Sulawesi to Enlarge a Picture**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

Appendix Figure 6.8 **Guidelines and Options to Make a Print Out**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

- Wait until PowerPoint appears. This opens an Microsoft (MS) PowerPoint application and retrieves a working file or loads a blank slide where the map can be affixed.
- Click right button of mouse and select Paste. This copies and pastes the map onto a PowerPoint slide.
- Adjust the size and layout. This corrects the size or crops the picture if needed.
- Print the layout from the PowerPoint menu. This prints the map.
- Save the file using another name (if necessary). This saves the file as a PowerPoint or graphic file.
- The clipboard can be pasted to another application by opening an application and pasting it. This allows the user to paste the picture on to the clipboard for use with other applications like MS Word.

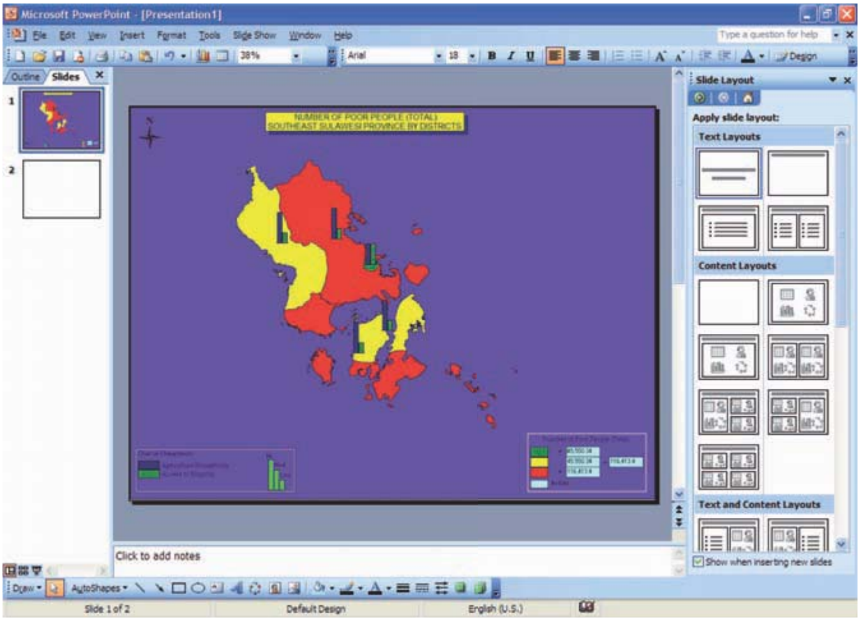
Using the Maps in Microsoft Applications

PRISMA allows maps to be used in MS applications using the processes described above or by using the computer's *Print Screen* function. Pressing the *Print Screen* (Prt Sc) key, copies the map currently on the screen to a clipboard from which the map can be copied (by going to Edit and selecting Paste) in MS PowerPoint, MS Word, and MS Excel. The maps can also be used with MS Publisher, MS Access, Paint, and WordPad.

Appendix Figure 6.9 shows the number of poor people in urban and rural areas in the districts Southeast Sulawesi, with an overlaid bar chart of the percentage of agriculture households and the percentage of villages with access to hospitals. The thematic map is transferred to the PowerPoint environment through the use of the print menu. Legends and the north-orientation sign are included. The figure shows that above-average poverty incidence is particularly observed in the eastern and southern part of the province. These areas have a high percentage of households whose heads' primary occupation is agriculture, showing a positive association with poverty. In addition, these areas, as well as those with average occurrence of poverty, have little access to hospitals. The only area where access to hospitals is not a major problem is the provincial capital, Kendari, where the number of poor is below average.

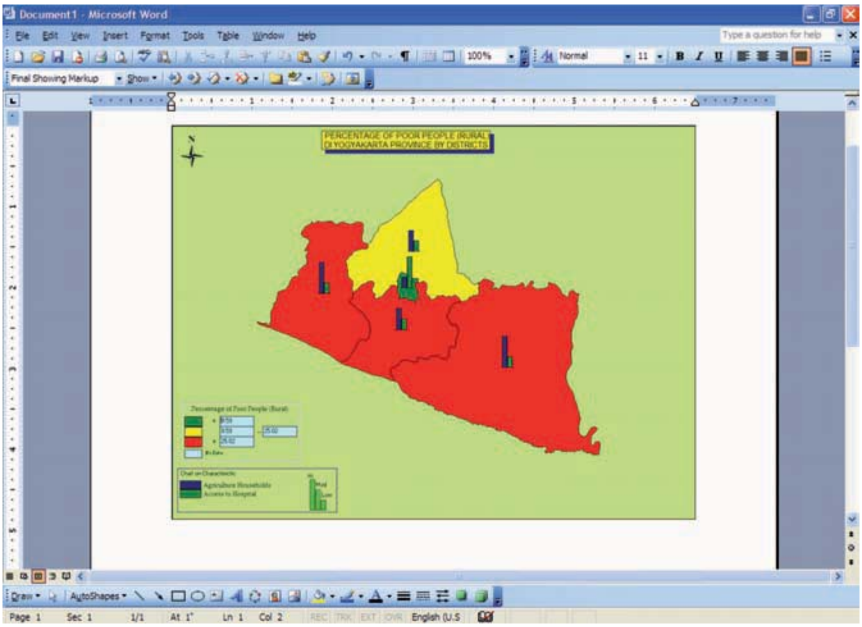
Appendix Figure 6.10 shows the percentage of poor people in rural areas in the districts of Yogyakarta. The map is also overlaid with the poverty characteristics of agricultural households and access to hospitals and is pasted as a picture on a Word document. The map shows high incidence of poverty throughout the province. Agricultural households are also prevalent in these areas and access to hospitals is a major consideration in these poor areas. The background of the picture has been altered and the legends moved to the lower left of the map to improve the presentation of this information.

Appendix Figure 6.9 **Exportation of a map from PRISMA to Microsoft PowerPoint**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

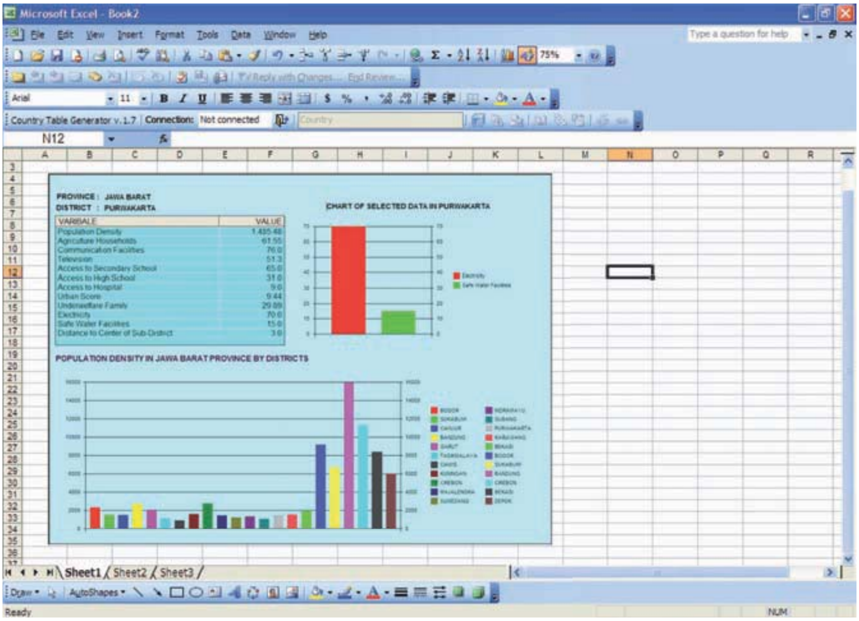
Appendix Figure 6.10 **Exportation of a Map from PRISMA to Microsoft Word**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.

Maps and information charts can also be used in MS Excel. For example, Appendix Figure 6.11 is a map in Excel that contains an information table and charts pertaining to the district of Purwakarta in the province of Jawa Barat (West Java). The bar chart on the table's right shows that, in Purwakarta, a high percentage of households have access to electricity, but a low percentage have access to safe-water facilities. The bar chart below the table shows that the district is among those with the least dense population in West Java; the highest is Bandung, followed by Cirebon.

Appendix Figure 6.11 **Exportation of the Information Charts from PRISMA to Microsoft Excel**



Source: Poverty Reduction Information System for Monitoring and Analysis (PRISMA), 2005.