SMERU

Working Paper

Sudarno Sumarto

Daniel Suryadarma

Asep Suryahadi

Predicting Consumption Poverty Using Non-consumption Indicators: Experiments Using Indonesian Data

February 2006

For further information, please contact SMERU, Phone: 62-21-31936336; Fax: 62-21-31930850; E-mail: smeru@smeru.or.id; Web: www.smeru.or.id

Predicting Consumption Poverty Using Non-consumption Indicators: Experiments Using Indonesian Data

Sudarno Sumarto

Daniel Suryadarma

Asep Suryahadi

SMERU Research Institute

February 2006

Predicting Consumption Poverty Using Non-consumption Indicators: Experiments Using Indonesian Data/
Sudarno Sumarto, Daniel Suryadarma, and Asep Suryahadi. –
Jakarta: SMERU Research Institute, 2006. ii, 29 p.; 31 cm. –(SMERU Working Paper, February 2006). –

ISBN 979-3972-20-9

1.Poverty Measurement

I. Sumarto, Sudarno
II. Suryadarma, Asep
III. Suryahadi, Asep

362.57/DDC 21

SMERU Research Institute, February 2006

TABLE OF CONTENTS

Predicting Consumption Poverty Using Non-consumption Indicators: Experiments Using Indonesian Data

Sudarno Sumarto
Daniel Suryadarma
Asep Suryahadi

SMERU Research Institute

February 2006

ABSTRACT

Although consumption expenditure data is crucial for assessing the level of people's welfare and calculating important welfare measures such as the poverty headcount rate, collecting such data requires significant time and effort. In this study, we experiment with three approaches to predict consumption expenditure and poverty at household and aggregate level as simpler alternatives to using consumption expenditure. The idea is not to use these alternatives as a substitute for consumption expenditure data, rather to use it for the purposes of rapid monitoring and appraisal of welfare. The three approaches are i) consumption correlates model, ii) poverty probability model, and iii) the wealth index Principal Components Analysis (PCA). We test each approach's performance and found that the consumption correlates model is the best approach to predict poverty quickly and relatively accurately. We found that education level, asset ownership, and consumption pattern are the best predictors of expenditure and poverty.

Keywords: consumption, poverty, predictor, data, Indonesia.

I. INTRODUCTION

According to the official statistics, in 2004 it was estimated that around 16.6 percent out of a total Indonesian population of around 210 million was considered to be living in absolute poverty. In addition, around one third to one half of the population was considered vulnerable to poverty, implying that they could easily fall into poverty when negative shocks occur.[1] Consequently, poverty reduction efforts will continue to be an important endeavor in Indonesia, even long into the future.

An important part of this endeavor is poverty monitoring, which is a regular and systematic effort to monitor welfare condition of the society. An effective poverty monitoring will be able to detect signs of deterioration in people's welfare, so that the relevant authorities can take appropriate and timely action to prevent the condition from worsening. Such monitoring requires regular and frequent estimations of various welfare measurements or indicators. Hence, a regular and frequent effort to collect household level data is a crucial part of an effective poverty monitoring system.

The most widely used data for measuring welfare or poverty is household consumption expenditure, especially in developing countries where household income data is considered more difficult to collect and less accurate. Another advantage of using expenditure data is that it is already expressed in monetary units, which is very intuitive and easily understood. So, one can proceed to calculate a poverty line using one's choice of method, and then simply calculate the poverty headcount rate, i.e. the proportion of people living below the poverty line.

In reality, however, it is not quite as simple. Collecting household consumption expenditure data requires a lot of time and effort. Firstly, it requires a sufficiently large dose of patience and willingness on the part of respondents, since they usually have to self-document their expenditure for a period of time, and sufficient trust on the part of the enumerator to entrust the respondents to correctly record their actual expenditure. Secondly, it takes a long time, about one week in Indonesia to record a pattern of food expenditure. Thirdly, if the questions involve the memory of respondents in recalling their non-food expenditure over the last 12 months, then reliability is an important issue.

In the face of these difficulties, a number of studies in developing countries have tried to address this empirical problem by creating a proxy for expenditure or poverty. The proxy is calculated using several widely recognized methodologies using easier to collect household characteristics data that have been proven to significantly influence poverty, such as assets ownership and education level. The main purpose is to get a figure that ranks households to the same place they would be had they been ranked by per capita consumption expenditure.
One of the more widely cited studies on estimating household expenditure is Filmer & Pritchett (2001), which use the Principal Components Analysis (PCA) method to calculate long-term household wealth in India and uses it as an explanatory variable of school enrolment. Meanwhile, Ward *et al.* (2002) and Abeyasekera *et al.* (2002) use the Ordinary Least Squares (OLS) regression method on a survey data from Tanzania to predict expenditure and income poverty. A similar study in Africa is Geda *et al.* (2001), which uses

[1] Pritchett *et al.* (2000), Suryahadi & Sumarto (2003).

data from Kenya. To test the performance of their models in predicting welfare, these studies basically compare the rank of households using the new index with the rank based on expenditure.

Since expenditure data is often used directly to measure poverty, several studies use the PCA (Principal Components Analysis) or MCA (Multiple Correspondence Analysis) method not to estimate a proxy for expenditure, but directly to measure poverty on a broader scope: multidimensional poverty. In a nutshell, this concept argues that poverty does not only involve expenditure or income dimension, but also other dimensions such as health, education, social status, and leisure.[2] Studies that adopt this approach include Asselin (2002) and Reyes *et al.* (2004).

In Indonesia, the only government agency known to collect non-expenditure data nationally to determine household welfare is BKKBN (*Badan Koordinasi Keluarga Berencana Nasional*, the National Family Planning Coordination Board). BKKBN uses 23 indicators to classify a household into five categories: (i) pre-prosperous family; (ii) just prosperous family; (iii) prosperous level 2 family; (iv) prosperous level 3 family; and (v) prosperous level 3 plus family. BKKBN has been collecting the data nationally every year since 1994, although due to the decentralization in Indonesia that transferred family planning matters to district governments, currently only 99 out of 400 districts in Indonesia still conduct the census.[3]

In this paper we try to predict expenditure and poverty in Indonesia using various household characteristics, such as asset ownership, housing condition, access to facilities, consumption pattern, as well as household head and spouse characteristics such as age, gender, and education level. To our knowledge, this is the second paper after Filmer & Pritchett (2001) that attempts to calculate a proxy for household expenditure or poverty in Indonesia.

The purpose of this study is to obtain relatively accurate indicators of poverty, both at household and aggregate levels, which are relatively easy and quick to collect and do not require household income or expenditure data. Therefore, it is important to bear in mind that we do not intend to estimate a full model of the determinants of poverty in Indonesia. Since there is no single agreed method to identify these variables, we explore three approaches – estimating a correlates model of consumption, estimating a limited dependent variable model of poverty, and calculating a wealth index – which have all been used in similar analyses in other countries, and compare the performance of each approach with the actual data in predicting poverty empirically.[4] The rest of the paper is organized as follows: section II discusses the methods and data, sections III-V discuss the results and performance of each method, and section VI concludes.

[2] Since this study is not about multidimensional poverty, interested readers should consult other studies that specifically address this issue. Recent publications include Duclos *et al.* (2001); Costa (2002, 2003); Tsui (2002); and Bibi (2004).

[3] Sumarto *et al.* (2004) discusses the BKKBN data in detail, including its compatibility with expenditure data.

[4] This is also the path taken in Filmer & Pritchett (2001), Geda *et al.* (2001), and Ward *et al.* (2002).

II. DATA AND METHOD

A. DATA

We use Susenas (the National Socioeconomic Survey) data in this study. Susenas is a nationally representative household survey, and has two main components. The first one is "Core" Susenas. It is conducted annually and collects data on household general characteristics and demographic information. The second component is "Module" Susenas, which contains more detailed characteristics of the households. There are three regular modules: (i) consumption; (ii) health, education, and housing; and (iii) social, crime, and tourism. They are conducted in alternating years, which means each module is conducted every three years.

The Core Susenas covers about 200,000 households – around 800,000 individuals – while the Module covers about 65,000 households that are randomly chosen from the Core sample. For this study we use Susenas 1999 because – in addition to the Core and Consumption Module – it has a special Social Safety Net (SSN) Module which records broader asset ownership and household characteristics information. Thus, the dataset we use is a merged dataset of the Core, Consumption Module, and SSN Module, with a total observation of 58,496 households.

B. METHOD

Based on a literature review, we have identified three methods that are commonly used in creating non-income/consumption poverty predictors: (i) estimating a correlates model of consumption, (ii) estimating a limited dependent variable model of poverty, and (iii) calculating a wealth index. We will explore all three methods in this study and, based on their performance, choose the most appropriate method to determine the poverty predictors to be used in the context of Indonesia. Furthermore, since it is widely recognized that conditions in urban and rural areas differ significantly in Indonesia, for each method we apply the method for urban and rural areas separately.

Method 1: Consumption Correlates Model

When poverty is defined as the current consumption deficit, a household is categorized as poor if the per capita consumption of its members is lower than the poverty line. Therefore, it is logical to search for poverty predictors based on variables that correlate with per capita household consumption. These variables can be obtained by estimating a correlates model of consumption, where the left hand side is per capita consumption and the right hand side is a set of variables that is thought to correlate with household consumption. In contrast to the determinants model, the endogeneity of the right hand side variables in the correlates model is not a concern.[5]

[5] Take for example the car ownership variable. Generally one would think that whether a household owns a car or not is determined by, among others, 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.

Once the set of right-hand side variables has been determined, a stepwise regression procedure is employed to estimate the model. The stepwise estimation procedure is used because in the end we want to obtain a manageable number of variables that can be relatively easily collected and at the same time meaningfully used to predict household consumption level and poverty status.

After the final set of variables has been determined, and their correlations with household per capita consumption have been estimated, the performance of this set of variables in predicting poverty can be empirically tested. First, the variables are used to predict per capita consumption level of all households in the sample. Second, the predicted per capita consumption is compared with the poverty line to determine the predicted 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. A similar test is also conducted to test the reliability of the model in predicting hardcore poverty, which is here defined as per capita household consumption below the food poverty line.

Method 2: Poverty Probability Model

The difference between this method and the first one is on the left hand side variable used in the estimation model. In the first method, the left hand side variable is per capita household consumption. In this method, on the other hand, household poverty status is directly used as the dependent variable. Since this is a binary variable, the model is a type of limited dependent variable model. This has an implication on the estimation procedure used, which here is based on the probit model.

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 of household poverty status. On the other hand, the use of the consumption correlates 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 consumption correlates 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 those who are already above the poverty line. Although positively correlated with welfare in general, such factors will not be correlated with poverty.

In this method, we use the probit estimation procedure where the right hand side variables are the same as in the consumption correlates model. Similarly, a stepwise estimation procedure is also used in order to come up with a manageable number of poverty predictors. As in the first method, a reliability test of these predictors to predict poverty is implemented by cross-tabulating the predicted with actual poverty status of households. Two levels of poverty – total and hardcore poverty – are also examined in this method.

Method 3: Wealth Index PCA

One relatively easily collected and reliable source of data on indicators of household socio-economic level is assets ownership. Using data on assets ownership, the wealth ranking of households can be ascertained through the creation of a wealth index. Unfortunately, data on assets ownership is usually only available in the form of binary variables, indicating only whether a household owns certain kinds of assets or not. Meanwhile, to create an appropriate wealth index requires data on the quality or price of each asset owned by a household to appropriately weigh each asset owned by a household. Hence, binary data poses a problem in ranking households by their socio-economic levels.

To deal with this problem, we utilize the Principal Components Analysis (PCA) method. In this method, the weighting for each asset is determined by the data itself. Intuitively, principal components 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.[6] The first principal-component is the linear index of variables with the largest amount of information common to all of the variables. 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). PCA, however, measures only relative poverty.

Based on the results of this analysis, households can be ranked from lowest to the highest socio-economic level. To test the reliability of this wealth ranking on predicting poverty requires a cutoff point to separate between the predicted poor and non-poor. Since there is no *a priori* poverty line that can be determined objectively, the cutoff point used is determined by the location of the poverty line in the actual consumption expenditure. In other words, the cutoff point is the proportion of poor households based on the actual consumption expenditure. As was the case in the previous two methods, a cross-tabulation is performed between the results of this approach and the poverty status based on the actual consumption expenditure. Also, as before, two levels of poverty – total and hardcore poverty – are examined in this approach.

The Poverty Line

The Indonesian poverty line and food poverty line used in this study are the ones calculated by Pradhan *et al.* (2001). This food poverty line is based on a single national bundle of food producing 2,100 calories per person per day priced by nominal regional prices. This means that the differences in the value of this food poverty line across regions arise solely due to price differences across regions. The non-food poverty line component, meanwhile, is estimated based on the Engel law method. The total and food poverty lines used in this study are shown in Table A1 in the Appendix.

[6] Filmer and Pritchett (2001).

C. VARIABLES USED

For all estimations, we use the same set of initial variables, which are narrowed down using stepwise regression in the first and second approaches. This means the final sets of variables for urban and rural areas for each method may differ.

In order to filter out multicollinearity, we first calculate the correlation coefficient of each variable within each group, drop some highly correlated variables, and then calculate the correlation coefficient with variables from other groups. The final set of variables is listed in Table A2 in the Appendix. These variables do not have a correlation coefficient higher than 0.7 with any other variables, implying that the multicollinearity issue has been minimized.

In method 1 we use real per capita household expenditure as the dependent variable. This is the nominal per capita expenditure deflated by the poverty line in the respondent household's province of residence in order to capture the price difference across regions in Indonesia. Thus, by using deflated per capita expenditure we have a measure of expenditure that is comparable across the country. On the other hand, in method 2 we use a binary variable that is 0 when the household is not poor and 1 when the household is poor. There are two measures of poverty that we examine: total poor and hardcore poor.

III. RESULTS OF CONSUMPTION CORRELATES MODEL ESTIMATIONS (METHOD 1)

As already mentioned above, in estimating the consumption correlates model, we estimate separate stepwise OLS regression for urban and rural areas. The final results of the estimations are shown in Table 1. Variables that were dropped from the model in both urban and rural areas by the stepwise procedures are unlisted in the table.

The models can explain 44% of variations in per capita consumption of urban households and 36% for rural households. This is similar to the result from Tanzania (Ward *et al.*, 2002) where around 40% of variance in expenditure can be explained. Furthermore, most of the coefficients have signs as expected. However, some variables have correlations with consumption only in urban areas but not in rural areas and *vice versa*. In addition, the coefficients of some variables have opposite signs in urban and rural areas.

Among the asset ownership group of variables, in urban areas all the coefficients are positive, indicating that owning these various assets is correlated with higher levels of household welfare. In both urban and rural areas, the ownership of cars, refrigerators, motorcycles, and satellite dishes are the variables with the highest correlation 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 correlated with lower per capita consumption.

Like assets ownership, the coefficients for house characteristics variables indicate that better housing materials are correlated with higher per capita consumption. In urban areas, tile roof and concrete wall are the two house characteristics that have the highest correlations with consumption. In rural areas, meanwhile, having electricity connection to the house and using flush toilets are the two house characteristics that have the highest correlations with consumption.

Table 1. Stepwise OLS Regression Results of Consumption Correlates Model (dependent variable: log of real per capita expenditure)

	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/s	-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**
house has dirt floor		-0.054**
		[0.008]
this household's source of water is from a protected well or water pump	0.078**	0.045**
house has flush toilet	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 has finished primary education	0.168**	0.030**
	[0.017]	[0.008]
household head has finished junior secondary education	0.245**	0.092**
	[0.022]	[0.019]
household head has finished senior secondary education	0.395**	0.150**
	[0.026]	[0.019]
household head has finished tertiary education	0.734**	0.292**
	[0.046]	[0.042]

	Urban Areas	Rural Areas
…Household Characteristics		
spouse has finished primary education	-0.123**	-0.038**
	[0.021]	[0.009]
spouse has finished junior secondary education	-0.178**	-0.051**
	[0.029]	[0.018]
spouse has finished 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]
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 of 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

Note: Robust standard errors in brackets.

* significant at 5%; ** significant at 1%.

The correlations of age with consumption also differ in urban and rural areas. In rural areas, it is the age of household head which matters and the relationship is positive at a decreasing rate. On the other hand, in urban areas, it is the age of the household head's spouse which matters and the relationship is negative at a decreasing rate.

Education level of 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.

On the other hand, the education level of spouse is negatively correlated with consumption. This is an unexpected and puzzling result, but the findings are consistent 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 correlation is controlled by holding other variables constant. One possibility is that these negative coefficients may indicate that, *ceteris paribus*, households with higher education levels of spouses save more, hence they consume less.

In rural areas, the enrollment status of school-age children also significantly correlates 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.

Both in urban and rural areas, larger household size is correlated with lower per capita consumption. The coefficients of household size square variable indicate that the lowering of 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.

Working status of 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. Then in both urban and rural areas, households which consume beef, egg, milk, biscuit, bread, and banana at least once per week have higher per capita consumption. On the other hand, in rural areas, households which consume *tiwul* (cassava flour), which is often considered 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 the actual consumption, where both the predicted and actual consumption

is ranked and divided into three groups: bottom 30%, middle 40%, and top 30%. Table 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% and off-diagonal cells will be 0.

Table 2. Accuracy of Predicted Expenditure Using Method 1

Row percentage of urban consumption expenditure

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

Row percentage of rural consumption expenditure

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

In urban areas, 67.3% of households in the bottom 30% are predicted correctly, while only 2.45 % of those households are wrongly predicted to be in the top 30%. Meanwhile, for those who are actually in the top 30%, 69.6% are predicted correctly, while about 2.7% are wrongly predicted to be in the bottom 30%. For the 40% in the middle, 56.6% are accurately predicted, while the rest 43% are predicted almost equally split to be in the top or bottom 30%.

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

On average, 64.5% of households' position in the per capita consumption groups is predicted correctly in urban areas and 60.8% in rural areas. As expected, the prediction in urban areas is more accurate because of the higher coefficient of determination in the regression results.
The next step is to examine the accuracy of the model in predicting poverty. Since poverty lines have already been established, we classify as poor those whose predicted expenditure is below the poverty line. Table 3 shows the result for poverty and Table 4 for hardcore poverty. Since our interest is in predicting poverty, the accuracy of predicting the non-poor is less relevant. As Table 3 shows, in urban areas around 49.6% of the poor are correctly predicted as poor, while the result is slightly lower in rural areas, where 45.7% are correctly predicted. This indicates that predicted expenditure tends to underestimate poverty. So if used as a targeting tool for the poor in urban areas, there will be an undercoverage of 50.4%, the share of poor who are wrongly predicted to be non-poor, and on the other hand about 7.3% of the non-poor will benefit from the program.

Table 3. Accuracy of Predicting Poverty Using Approach 1

Row percentage of urban poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	92.73	7.27
Poor	50.43	49.57

Row percentage of rural poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	92.12	7.88
Poor	54.32	45.68

Meanwhile, Table 4 shows that the prediction results are even lower for hardcore poverty. Around 48.4% of the hardcore poor in urban areas and only some 33.5% of the hardcore poor in rural areas are correctly classified.

Table 4. Accuracy of Predicting Hardcore Poverty Using Approach 1

Row percentage of Urban Poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	94.62	5.38
Poor	51.55	48.45

Row percentage of Rural Poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	95.60	4.40
Poor	66.52	33.48

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

IV. RESULTS OF POVERTY PROBABILITY MODEL ESTIMATIONS (METHOD 2)

This approach predicts poverty directly because of the nature of the dependent variable. 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 otherwise. 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 being poor. This is in contrast with Approach 1, where a positive coefficient means that the variable increases expenditure, hence reduces the chances of being poor. Second, the predicted value of the dependent variable is the probability of the observation to be poor. The result of estimation of poverty is in Table 5 while the result of the hardcore poverty estimation is in Table 6.

For the poverty estimations, the pseudo R-squared is 0.36 for urban areas and 0.29 for rural areas. Meanwhile, for hardcore poverty estimations, the pseudo R-squared is 0.35 for urban areas and 0.28 for rural areas. In general, the coefficients in Table 5 are consistent with those in Table 1. For example, the asset ownership variables have positive coefficients in Table 1, which means that households that own various assets are more likely to have higher consumption expenditures. In Table 5, meanwhile, 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.

However, there are some exceptions. For example, in Table 1 the variable of owning a sewing machine is dropped in the process 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 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 6 has fewer significant variables – the insignificant ones are already dropped by the stepwise regression – which mean that there are fewer hardcore poverty predictors than poverty predictors. For example, in regard to the education level of household head, the results in Table 5 indicate that, relative to households with heads having education less than primary level, the higher the education level of household head, the lower the probability of that household to be poor. However, the results in Table 6 indicate that only households whose heads have at least a senior high school level education have a significantly lower probability of being hardcore poor.

Table 5. Stepwise Probit Estimation Results of Poverty Probability Model (dependent variable: 1 = poor, 0 = otherwise)

	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 cow/s		0.065**
		[0.022]
this household owns chicken/s		-0.106**
		[0.017]
this household owns other animal/s	0.403**	
	[0.141]	
House Characteristics		
wall of the house is made from concrete	-0.206**	-0.137**
house has dirt floor	0.214**	0.144**
	[0.049]	[0.023]
house has a flush toilet	-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 water source is a 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 has finished primary education	-0.111**	-0.082**
	[0.034]	[0.021]
household head has finished junior secondary education	-0.210**	-0.134**
	[0.043]	[0.034]

	Urban Areas	Rural Areas
… Household Characteristics		
household head has finished senior secondary education	-0.271**	-0.245**
	[0.044]	[0.041]
household head has finished tertiary education	-0.640**	-0.517**
	[0.104]	[0.126]
spouse has finished primary education		0.087**
		[0.021]
household size	0.627**	0.649**
	[0.028]	[0.021]
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 of 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]
this household 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

Note: Robust standard errors in brackets.
* significant at 5%; ** significant at 1%.

Table 6. Stepwise Probit Estimation Results of Hardcore Poverty Probability Model (dependent variable: 1 = hardcore poor, 0 = otherwise)

	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 chicken/s		-0.101**
		[0.021]
House Characteristics		
wall of the house is made from concrete	-0.286**	-0.166**
house has dirt floor		0.135**
house has flush toilet	-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 a 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 has finished senior secondary education	-0.283**	-0.165**
	[0.066]	[0.052]
household head has finished tertiary education	-0.960**	
	[0.287]	
spouse has finished primary education		0.066**
		[0.023]
household size	0.509**	0.590**
	[0.039]	[0.023]

	Urban Areas	Rural Areas
… Household Characteristics		
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]
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 of 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
Number of observations	23,759	34,649
Pseudo R-squared	0.352	0.28

Note: Robust standard errors in brackets.
* significant at 5%; ** significant at 1%.

We now turn to see the accuracy of predicting actual poverty using this second approach. The predicted value of the dependent variable in this approach is the probability of households to be poor given their characteristics. To classify households into predicted poor and predicted non-poor, we need a threshold to separate these two groups of households. Following Pritchett *et al.* (2000) and Suryahadi & Sumarto (2003), we use a 50% probability of being poor as the threshold. Hence, households which have fifty-fifty 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 non-poor. Using this 50% probability threshold, Tables 7 and 8 show the cross-tabulations between the actual and predicted incidence of respective poverty conditions.

Table 7. Accuracy of Predicting Poverty Using Approach 2

Row percentage of urban poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	97.07	2.93
Poor	64.44	35.56

Row percentage of rural poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	90.49	9.51
Poor	47.33	52.67

Table 7 shows that 35.6% of the poor are predicted correctly in urban areas and less than three percent of the non-poor are predicted to be poor. Meanwhile, in rural areas about 52.7% of the poor are predicted correctly, even though the percentage of the non-poor predicted to be poor is also higher, 9.5%.[7] Prediction for urban areas is much less accurate than using Approach 1, where almost 50% of the poor are correctly predicted. However, the prediction in rural areas is better than using Approach 1.

Table 8. Accuracy of Predicting Hardcore Poverty Using Approach 2

Row percentage of urban poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	99.66	0.34
Poor	87.89	12.11

Row percentage of rural poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	97.62	2.38
Poor	73.67	26.33

Meanwhile, Table 8 shows that the predicted hardcore poverty is even less accurate than the predicted poverty. Comparing Table 8 with Table 4, Approach 2 gives a worse prediction than Approach 1.

[7] Changing the 50% threshold of poverty probability will also change the accuracy. For example, by using 30% as the threshold, a higher rate of accuracy is achieved, however, the justification for choosing this threshold is hard to find. Thus a 50% threshold is considered to be a more logical threshold.

V. RESULTS OF WEALTH INDEX PCA (METHOD 3)

We follow Filmer and Pritchett (2001) in calculating the PCA index:[8]

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

where f_1 is the 'scoring factor' for the first asset determined by the method, a_{j1} is the j^{th} household's value for the first asset and a_1 and s_1 are the mean and standard deviation of the first asset variable over all households. Table 9 provides the scoring factor, mean, and standard deviation of each variable for urban areas, while Table 10 provides the information for rural areas. Mean of the indexes in both areas are zero by construction.

Table 9. Eigenvalue 1 (First Principal Component) and Summary Statistics for Urban Areas

	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
house has flush toilet	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 water source is a protected well or water pump	0.115	0.867	0.340	0.34
this household owns cow/s	-0.055	0.019	0.137	-0.40
this household owns goat/s	-0.048	0.019	0.135	-0.35
this household owns chicken/s	-0.053	0.152	0.359	-0.15
this household owns other animal/s	-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 has finished primary education	-0.105	0.247	0.431	-0.24
household head has finished junior secondary education	-0.005	0.165	0.371	-0.01
household head has finished senior secondary education	0.138	0.290	0.454	0.30
household head has finished tertiary education	0.180	0.097	0.297	0.61
spouse has finished primary education	-0.050	0.240	0.427	-0.12

[8] They refer to it as Economic Status Index. Although Filmer & Pritchett (2001) cautioned that they are not proposing the wealth index to 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.

	Scoring Factor	Mean	Standard Deviation	Scoring Factor/ Std Dev
spouse has finished junior secondary education	0.055	0.144	0.351	0.16
spouse has finished senior secondary education	0.184	0.194	0.395	0.47
spouse has finished 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 of this household is sick, s/he is treated with modern medicine	0.091	0.926	0.262	0.35
this household consumed *gaplek* (dried cassava) 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	

Table 10. Eigenvalue 1 (First Principal Component) and Summary Statistics for Rural Areas

	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
house has flush toilet	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 water source is a protected well or water pump	0.168	0.504	0.500	0.34
this household owns cow/s	-0.066	0.179	0.384	-0.17
this household owns goat/s	-0.049	0.114	0.318	-0.16
this household owns chicken/s	-0.035	0.465	0.499	-0.07
this household owns other animal/s	-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 has finished primary education	-0.003	0.339	0.474	-0.01
household head has finished junior secondary education	0.073	0.094	0.292	0.25
household head has finished senior secondary education	0.185	0.095	0.293	0.63
household head has finished tertiary education	0.140	0.019	0.136	1.03
spouse has finished primary education	0.039	0.300	0.458	0.09
spouse has finished junior secondary education	0.099	0.072	0.258	0.38
spouse has finished senior secondary education	0.170	0.055	0.228	0.75
spouse has finished 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 of this household is sick, s/he is treated with modern medicine	0.108	0.892	0.311	0.35

	Scoring Factor	Mean	Standard Deviation	Scoring Factor/ Std.Dev
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

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 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 decreases the index by 0.45. In urban areas, the lowest index is -7.31 while the highest is 6.6, while for rural areas the lowest is -4.82 and the highest is 10.52.

Table 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% of those in the bottom 30% and 54.6% of those in the top 30% are predicted correctly using Approach 3. Meanwhile, in rural areas 47.4% of those in the bottom 30% and 50.3% of those in the top 30% are accurately predicted. The accuracy of this approach is much lower compared to Approach 1, where more than 60% of each tercile is predicted correctly.

Table 11. Accuracy of Predicted Per Capita Consumption Expenditure Using Approach 3

Row percentage of urban consumption expenditure

	Predicted Based on Wealth Index		
Actual	**Bottom 30%**	**Middle 40%**	**Top 30%**
Bottom 30%	51.10	41.52	7.38
Middle 40%	25.79	45.69	28.52
Top 30%	14.51	30.89	54.61

Row percentage of rural consumption expenditure

	Predicted Based on Wealth Index		
Actual	**Bottom 30%**	**Middle 40%**	**Top 30%**
Bottom 30%	47.35	40.73	11.92
Middle 40%	26.84	44.78	28.38
Top 30%	16.85	32.90	50.25

To measure the performance of this approach in predicting poverty, we need to have a threshold to divide households into those that are predicted as poor and non-poor. Since there is no such threshold in the wealth index that can be calculated objectively, we assume that the threshold is the value of the wealth index at the percentile of the actual poverty rate. For example, if the poverty rate is X%, 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% predicted poverty rate, which is the same as the actual poverty rates. Using this threshold, Tables 12 and 13 show the cross-tabulation between the actual and predicted rates for poverty and hardcore poverty respectively.

Table 12. Accuracy of Predicting Poverty Using Approach 3

Row percentage of urban poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	90.14	9.86
Poor	64.72	35.28

Row percentage of rural poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	78.12	21.88
Poor	53.68	46.32

Table 12 shows that only 35.3% of the poor in urban areas are predicted correctly, making wealth index PCA the least accurate approach for predicting poverty compared to the other two approaches for urban areas. However, 46.3% of people in rural areas are predicted correctly, which is slightly higher than using Approach 1 – which is 45.7% – although lower than Approach 2 – which is 52.7%.

Table 13. Accuracy of Predicting Hardcore Poverty Using Approach 3

Row percentage of urban poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	96.43	3.57
Poor	81.68	18.32

Row percentage of rural poverty

	Predicted	
Actual	**Not Poor**	**Poor**
Not Poor	89.20	10.80
Poor	68.14	31.86

Meanwhile, in predicting hardcore poverty, 18.3% of the hardcore poor in urban areas and 31.9% in rural areas are predicted correctly. Compared to the performance of the other approaches in predicting hardcore poverty, the accuracy of this approach is higher than Approach 2 but lower than Approach 1.

VI. CONCLUSION

In the face of the difficulties in acquiring household expenditure and income data, and to find variables that predict poverty the best, in this study we explore three approaches on predicting poverty. These three approaches are the consumption correlates model, poverty probability model, and wealth index Principal Component Analysis (PCA). In terms of predicting expenditure, the consumption correlates model is the best approach as it is able to predict correctly more than 60% of the poor respondents in both urban and rural areas, while the accuracy of the wealth index PCA on average is less than 50%.

Meanwhile, in terms of predicting poverty and hardcore poverty, the results are mixed. A common thread in the predictions is that the better the poverty prediction is, the more non-poor is predicted to be poor. So it is a double-edged sword because on one hand we want the method that gives us the most accurate prediction, but this method will also predict the most non-poor to be poor. However, since we believe that the focus of this paper is to predict the poor, we should focus on the methodology that provides us with the most accurate prediction on the poor. In hardcore poverty prediction, by far the best approach is the consumption correlates model. In predicting poverty, the poverty probability model is the best predictor for rural areas – 52.7% accurate – while for urban areas the consumption correlates model gives the best result – 49.6% accurate. In conclusion, we believe that the consumption model is, on balance, the best approach to use to find expenditure and poverty predictors.

The variables with the strongest correlates, either negative or positive, are car and refrigerator ownership; education level; household size; and consumption of milk and beef. Furthermore, house characteristics, access to facilities and employment status of household members play relatively small but significant roles. Thus, if we want to assess whether a household is more likely to be poor or not in Indonesia, it would be better to gather information on asset ownership, education level, and consumption pattern.

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 it is postulated that quality also makes much of the difference between poor and non-poor households.

APPENDIX

Table A1. Poverty Lines in February 1999
(Rp per capita/month)

	Poverty Line		Food Poverty Line	
Province	**Urban**	**Rural**	**Urban**	**Rural**
Aceh	74,064	70,280	60,733	60,003
North Sumatra	83,745	74,712	66,803	63,753
West Sumatra	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 Sumatra	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	-	76,747	-
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

Source: Pradhan *et al*. (2001).

Table A2. Variable List 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're living in
House	tileroof	roof of the house is made from tile
	concretewall	wall of the house is made from concrete
	dirtfloor	house has dirt floor
	flushtoilet	house has flush toilet
	own_toilet	this household uses its own toilet
	electric_light	this household has electricity
	protectedwatersrc	this household's source of water is from a protected well or water pump
Farm	own_cow	this household owns cow/s
	own_goat	this household owns goat/s
	own_chicks	this household owns chicken/s
	own_othanim	this household owns other animal/s
Household	age	household head age
	spage	spouse age
	elm	household head has finished primary education
	lsec	household head has finished junior secondary education
	usec	household head has finished senior secondary education
	ter	household head has finished tertiary education
	spelm	spouse has finished primary education
	splsec	spouse has finished junior secondary education
	spusec	spouse has finished senior secondary education
	spter	spouse has finished 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

Group	Variable	Description
	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 of 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.

REFERENCES

Abeyasekera, Savitri and Patrick Ward (2002), 'Models for Predicting Expenditure per Adult Equivalent', Tanzanian Ministry of Health, UK DFID, University of Newcastle upon Tyne, and Districts of Hai, Ilala, Morogoro, Rufiji, Temeke, Adult Morbidity and Mortality Project, mimeo.

Asselin, Louis-Marie (2002), 'Composite Indicator or Multidimensional Poverty', CECI, Canada, mimeo.

Bibi, Sami (2004), 'Comparing Multidimensional Poverty between Egypt and Tunisia', Paper presented at conference of The Center for the Study of African Economies, University of Oxford, 21st to 22nd March 2004.

Costa, Michele (2002), 'A Multidimensional Approach to the Measurement of Poverty', IRISS Working Paper Series No. 2002-05, CEPS/INSTEAD, Luxembourg.

Costa, Michele (2003), 'A Comparison between Unidimensional and Multidimensional Approaches to the Measurement of Poverty', IRISS Working Paper Series No. 2003-02, CEPS/INSTEAD, Luxembourg.

Duclos, Jean-Yves, David Sahn, and Stephen D. Younger (2001), 'Robust Multidimensional Poverty Comparisons', Universite Laval, Canada, mimeo.

Filmer, Deon and Lant Pritchett (2001), 'Estimating Wealth Effects without Expenditure Data -- or Tears: An Application to Educational Enrollments in States of India', *Demography*, 38(1), pp. 115-132.

Geda, Alemayehu, Niek de Jong, Germano Mwabu, and Mwangi S. Kimenyi (2001), 'Determinants of Poverty in Kenya: A Household Level Analysis', Institute of Social Studies and Kenya Institute for Public Policy Research and Analysis, mimeo.

Pradhan, Menno, Asep Suryahadi, Sudarno Sumarto, and Lant Pritchett (2001), 'Eating Like Which 'Joneses'? An Iterative Solution to the Choice of Poverty Line Reference Group', *The Review of Income and Wealth*, 47(4), pp. 473-487.

Pritchett, Lant, Asep Suryahadi, and Sudarno Sumarto (2000), 'Quantifying Vulnerability to Poverty: A Proposed Measure, with Application to Indonesia', *SMERU Working Paper*, May, SMERU Research Institute, Jakarta.

Reyes, Celia, Kenneth Ilarde, Lani Valencia, and Joel Bancolita (2004), 'Utilizing CBMS in Monitoring and Targeting the Poor: The Case of Barangay, San Vicente, Palawan Unity-Based Monitoring System', Paper presented at PEP Research Network Meeting 2004, Senegal.

Setel, Philip, Savitri Abeyasekera, Patrick Ward, Yusuf Hemed, David Whitting, Robert Mswia, Manos Antoninis, and Henry Kitange (2003), 'Development, Validation, and Performance of a Rapid Consumption Expenditure Proxy for Measuring Income Poverty in Tanzania: Experience from AMMP Demographic Surveillance Sites', Paper prepared for the Asset Index Seminar, 26-28 March 2003, DFID's Health Systems Resource Centre and the University of Southampton.

Suryahadi, Asep and Sudarno Sumarto (2003), 'Poverty and Vulnerability in Indonesia Before and After the Economic Crisis', *Asian Economic Journal*, 17(1), pp. 45-64.

Tsui, Kai-yuen (2002), 'Multidimensional Poverty Indices', *Social Choice and Welfare*, 19(1), pp. 69-93.

Ward, Patrick, Trudy Owens, and Godius Kahyrara (2002), 'Developing Proxy Predictors for Household Expenditure and Income Poverty', Oxford Policy Management, mimeo.

Zeller, Manfred (2004), 'Review of Poverty Assessment Tools', Accelerated Microenterprise Advancement Project Report, United States Agency for International Development, Washington, DC.