

**Asian Development Bank**  
**People's Republic of China**

TA 4454 – Developing a Poverty Monitoring System at the County Level

# **County poverty-related indicators**

Report ©

Ludovico Carraro

**June 2005**

The views expressed in this paper are the views of the author and do not necessarily reflect the views or policies of the Asian Development Bank (ADB), or its Board of Directors or the governments they represent. ADB makes no representation concerning and does not guarantee the source, originality, accuracy, completeness or reliability of any statement, information, data, finding, interpretation, advice, opinion, or view presented.

## Background

In my previous report I made an assessment of the current poverty monitoring system in Dayao county as well as identified the main data needs for decision making. It emerged that there are county data needs at different levels of decision making (within the county and at the prefecture/province/nation) and that information is needed with a different degree of detail (household, village, and county). These data needs can be classified in three groups:

- 1) Information at the village level (administrative village and natural village): such information is required by the local government to decide about interventions/projects that can benefit the whole village (building roads, irrigation systems etc.);
- 2) Information at the household level: such information is needed by the local government for interventions at the village level, but also for projects directly targeted to individual households;
- 3) Information at the county level for upper government (prefecture, province and national governments): poverty indicators, especially the head-count, are required for a better allocation of poverty funds.

In designing a poverty monitoring system it is fundamental to prioritise data needs, making sure that collected data will be used, and that data collection will be feasible and cost-effective. Therefore, it is essential not only to identify appropriate indicators, but also to reduce the indicators to a manageable number.

We looked at three main types of possible indicators: 1) indicators highly correlated to income poverty (consumption expenditure); 2) indicators that are related to income poverty, but on their own are also capturing other important dimensions of living standards; 3) indicators relevant for local decision making.

In order to identify such indicators we used two different methodologies: 1) participatory approach with the key stakeholders in Dayao county (the results of such exercise were illustrated in the previous report); 2) regression analysis using household survey data (the 2004 China Rural Poverty Monitoring Survey for Yunnan province).

The first part of this report discusses in some detail the results of the survey analysis, how they can be used, and how the two methodologies are combined to identify a proposed list of indicators. A second section presents the list of indicators after taking into account the comments received in a second consultation with Dayao's government. A final section looks at the next steps and makes some general recommendations.

## Survey-based poverty indicators

### *The exercise*

Measuring income poverty, calculated through household consumption expenditure or income, involves expensive and long interviews with the household. In China there are two main sources of information of this kind: the Rural Household Survey (RHS) and the Rural Poverty Monitoring Survey (RPMS). However, such instruments cannot provide reliable estimates at the county level because the current sample size is too small, but the required increase in sample size would make such an exercise too expensive and not sustainable. Since it is not feasible to measure directly income poverty, we can monitor poverty indirectly, looking at indicators that are highly correlated to

income poverty, but much easier to collect. These indicators can then be used on their own, because they reveal different aspects of the household and community characteristics, but also jointly to predict poverty indicators (poverty head-count) for a relevant group of households (townships/county). At this stage we are mainly interested in the first use of the indicators, but I will also discuss the possible benefits of the second use.

Regression analysis using survey data have two main advantages: 1) testing the joint significance of the indicators, rather than the correlation of each indicator one by one; 2) identifying more accurately the importance of the various indicators, so that we can select a cost-effective number of indicators.

However, in testing and identifying the importance of the indicators we are limited by the information available through the household survey<sup>1</sup>. Moreover, such analysis might ignore some indicators that might be county specific.

### **Main results**

Using 22 indicators (variables) the regression explains about 50% of the variation in expenditure per capita<sup>2</sup>. This outcome is similar to results obtained by other researchers using the Rural Household Survey or the Rural Poverty Monitoring Survey. These variables include demographic characteristics of the household (household size and composition) education and health of household members, income sources, housing quality, ownership of consumer durable items and livestock, village location, presence of electricity and impact of natural disasters (the full results of such model are reported in the appendix). Although the main goal of this regression is to be able to predict as well as possible consumption levels, ignoring problems of endogeneity, most of the sign of the coefficients are in the expected direction<sup>3</sup>.

11 of these 22 most important indicators were also ranked among the most relevant in the assessment made by the government stakeholders. However, some of the remaining indicators were not considered at all by the participatory assessment (for instance demographic characteristics of the household) and in other cases the participatory assessment did not consider them very important (for instance the fact that the household head cannot speak Chinese). Finally there are indicators that were considered very important by the participatory assessment, but are not among the 22 most important indicators of the regression model. There are two main factors that explain such difference between the two methodologies: a) some of the indicators though highly correlated with poverty on their own, may not be important when considered together with other indicators, so that the regression analysis will omit such indicators<sup>4</sup>, b) some of the indicators identified by the Dayao stakeholders do not have direct equivalent or do not exist in the survey data (for instance rainfall).

---

<sup>1</sup> Theoretically, it would also be possible to use other sources of information at the village level, but this would require a considerable effort in combining administrative and survey data.

<sup>2</sup> More precisely the R<sup>2</sup> is 0.464. Using about 90 indicators the R<sup>2</sup> was only marginally higher at 0.495.

<sup>3</sup> A significant exception is the fact that natural disasters are positively correlated to consumption expenditure. As it will be discussed later, one possible explanation for this result is the fact that consumption expenditure includes also money spent for home improvements that might be more frequent after some types of natural disasters.

<sup>4</sup> Let's consider a hypothetical case. Both good housing and safe drinking water are highly correlated with consumption expenditure (poverty status), but all households living in good houses may also have access to safe drinking water, so that just one of the two indicators is required.

The proposed list of indicators will definitely need to include those that were highly important in both assessments, but it will also consider indicators that, although not among the 22 most important indicators, are considered very important by the county stakeholders and describe important dimensions of living standards. For instance, although safe water when jointly included with other indicators is not among one of the most important indicators, it nevertheless reveals one fundamental aspect of the quality of living. Moreover, other indicators will also be included even though they are not identified by the regression, because they are directly related to projects supported by policies of poverty reduction, and therefore they can provide relevant information for decision making (for instance presence and quality of motorable roads).

### ***Methodological aspects***

It is important to clarify some technical aspects related to the methodology adopted in the econometric modelling. These clarifications concern the use of weights, and methodological choices that should be made to eventually use such results to predict poverty levels.

#### 1) The use of population weights

The sample is made of 7300 households (in each of the 73 key poor counties the sample is made of 10 villages, and in each village 10 households are interviewed). In each county the villages are chosen from each income population decile and therefore within the county the sample is self-weighting. However, there are very big differences in the population size of the 73 counties and it would be inappropriate to consider that the sample as a whole is self-weighting. Therefore, I used a simple relative weight (county population divided by the sample population of each county)<sup>5</sup>. The use of the weights in the regression is important and this is confirmed by the canonical Hausman test.

#### 2) Use of the econometric results to predict poverty indicators

If the variables identified in the regression model are properly collected it is possible to use them to predict household consumption levels and therefore poverty levels. However, such prediction cannot be trustable for each individual household, but it would be possible to estimate with precision poverty indicators for a group of households. For instance, it would be possible to estimate poverty indicators at the county level and perhaps at the township level. Moreover, it could also be used to assess the targeting of government projects (poverty levels of households who receive some form of support from the government).

There are two different ways in which the computation of such indexes could be done.

a) The simpler way would be to test whether we can accept the hypothesis that the error term is normally distributed, and if such hypothesis is not refused the proportion of poor people in a certain population sub-group could be estimated as follows:

$$P_0 = \frac{1}{N} \sum \Phi \left( \frac{\ln z - X\beta}{\sigma} \right)$$

---

<sup>5</sup> This is an approximate post-sampling weight, but such procedure to compute them should be checked with the sampling experts of RSO to understand which sample frame was actually used in the selection of the villages and that their calculation is correct.

where  $\Phi$  is the cumulative function of the standard normal distribution,  $z$  is the poverty line,  $X\beta$  is the predicted consumption level and  $\sigma$  is the standard error of the regression (see Hentschel, J Lanjouw, J Lanjouw P and Poggi J (2000) "Combining census and survey data to study spatial dimensions of poverty: a case study of Ecuador", The World Bank Economic Review, Vol. 14, No. 1, pp. 147-165).

b) Since often the error term is not normally distributed, but affected by both cluster effect (autocorrelation) and heteroskedasticity, it would be possible to use the poverty mapping methodology developed by Elbers et al. (2001) to properly estimate consumption levels and poverty indicators (Elbers, C Lanjouw, J and Lanjouw P (2002): Micro-Level Estimation of Welfare, WB Policy Research Working Paper 2911).

The methods described above could be used if the new poverty monitoring system will be able to collect information for each village and household. Alternatively, if the collection of such data for each household will be ruled out because not feasible, such methodology could also be used if the same information will be collected for a representative sample of households in the county. In such case what would be required is an estimate of poverty from survey (RPMS or RHS) to survey (the new questionnaire developed for the county poverty monitoring system). However, this difference poses new methodological challenges, in particular on how to impute consumption in a way that takes into account the sampling error that comes from the new survey instrument. Although there are research papers that have tackled the issue of welfare estimation from sample to sample, such challenges are not fully resolved and this type of exercises is on the forefront of research<sup>6</sup>.

### ***Possible improvements in the econometric modelling***

There are various ways in which the current econometric results could be improved and further tested, and here I propose in some detail aspects that could be taken into consideration by the RSO in order to achieve better results<sup>7</sup>.

#### **1) Definition of the dependent variable**

The dependent variable is the consumption expenditure per capita, and this is obtained dividing the household living expenditure (variable x287) by the household size. However, currently both the calculation of the numerator and the denominator could be improved. For both of them I used variables that were already computed by the RSO, but it would be very important to re-compute correctly such aggregates.

Consumption expenditure. The current definition of consumption expenditure seems to reflect the notion of consumption expenditure used for the National Accounts rather than a definition appropriate for welfare analysis. The current consumption aggregate includes expenses that create a lot of noise and do not appropriately reflect living standards<sup>8</sup>: expenditure for durable goods, expenditures for home improvements, expenditures for important ceremonies (for example weddings) and exceptional health expenditures. All these expenditures can generally be very

---

<sup>6</sup> OPM conducted such sample to sample estimation in Bangladesh, and it is also discussing this type of approach in Tanzania.

<sup>7</sup> I also left at the RSO the STATA programs that I have written to make the regression analysis (files assembling, creation and definition of variables, methodology used to identify the most important variables, etc). As far as I could I inserted comments explaining what each part of the syntax does.

<sup>8</sup> I was surprised in finding that the variance of consumption expenditure was higher than the variance of income.

substantial, but they are unlikely to reflect the consumption patterns of that household (they are not made every year, and in some cases they are investments that improve living standards over a number of years rather than in one single year). The calculation of consumption expenditure for welfare analysis must exclude such items<sup>9</sup> (see Deaton and Zaidi (2002): Guidelines for constructing consumption aggregates for welfare analysis, LSMS Working Paper 135, WB). In order to better estimate the relationship between consumption and our indicators I included a series of dummy variables that show whether the household had any of such exceptional events, but this is clearly a second best solution.

Household size: the definition of household size used by the RSO includes all household members even if for the whole year they worked and lived somewhere else. However, although income of such household members is likely to be very important for the household it is unlikely that consumption expenditure will include expenses for such members. I therefore corrected the household size excluding household members who worked outside for more than six months, but again this calculation was made indirectly using some variables constructed by the RSO. Instead it would be more accurate to look at such cases in the individual records of the household where the reason for being absent and the exact number of months in which people live outside are reported.

## 2) Inclusion of more variables in the initial regression

Due to lack of time the number of variable interactions and transformations that were tested in the regression was incomplete, and it would be advisable to investigate further the effect of such interactions and transformations (in particular household size could be put in logarithm, different variables that consider the education of household members could be explored, for instance looking at education of male and female members of the household, interactions between education and type of employment should be created as well as interactions between community and household variables).

Currently the definitions of some community variables is not entirely satisfactory (we only have information about presence of roads within the administrative villages – natural villages), but we do not have information on the quality of the road from the administrative village to the township. Also we have information on the presence of primary school, but more important would be to consider whether the primary school can offer the complete number of grades or not. Such variables are not in the current dataset, but it would be possible to add this information from other sources.

## 3) Further testing the significance of variables

Given the wealth of household survey data available for rural China, it is possible and advisable to check the robustness of the results obtained through the regression analysis using the same dataset for previous years. Moreover, it would also be possible to test the model using data from the Rural Household Survey.

# A proposed list of county poverty indicators

As explained earlier the list of proposed indicators combines the results of the regression analysis and those of the participatory assessment. Moreover the list of indicators reported below is also the result of the second consultation with the local stakeholders.

---

<sup>9</sup> I could not make such calculation because it requires a re-analysis of the items included in the consumption aggregate (expenditure items in the household diaries).

## ***Indicators at the household level***

### *Demographic indicators*

- 1) Household size
- 2) Whether is an extended household (3 generations)
- 3) Proportion of household members 6 years and under (0-6)
- 4) Proportion of household members 60 or above

### *Education*

- 5) Highest education achieved by household members in working age
- 6) Whether there is any household member attending university
- 7) Whether any children aged 7-15 are not in school
- 8) Whether household head cannot speak Chinese

### *Health*

- 9) Number of disabled household members
- 10) Whether, if any household member is ill, he/she can get medical attention in time
- 11) To what extent illnesses of household members in working age affected their ability to work

### *Insurance*

- 12) Whether any of the household members have a health insurance?
- 13) Whether any of the household members have pension insurance?

### *Sources of income*

- 14) Proportion of cultivated land area used for cash crops
- 15) Proportion of household members in working age employed more than 6 months in agriculture
- 16) Proportion of household members in working age working out of the village for more than 6 months
- 17) Proportion of household members in working age employed in TVEs

### *Livestock*

- 18) Number of pigs (currently fed)
- 19) Number of pigs slaughtered or sold last year
- 20) Number of big animals (cattle, horses, buffalos etc.)
- 21) Number of big animals slaughtered or sold last year
- 22) Whether in the last 12 months the household suffered any livestock losses

### *Housing*

- 23) Main construction material used for the house
- 24) Type of fuel used for cooking

### *Water and sanitation*

- 25) Source of drinking water
- 26) Type of toilet
- 27) Distance to water source

### *Assets*

- 28) Whether the household has colour TV
- 29) Whether the household has telephone
- 30) Whether the household has motorbike
- 31) Whether the household has truck/tractor
- 32) Whether the household has small tractor

### *Other*

- 33) Distance to centre of administrative village
- 34) Type of support received from the government in the last 12 months

## ***Indicators at the village level***

### *Location*

- 1) Altitude and whether located in hill, mountain, or plane

### *Infrastructure and services*

- 2) Whether there is a motorable road from village to township
- 3) Type of road
- 4) On average for how many months a year the road cannot be used?
- 5) Proportion of natural villages with motorable road?
- 6) Proportion of natural villages with electricity?
- 7) Distance to complete primary school
- 8) Distance to middle school
- 9) Distance to hospital

### *Health<sup>10</sup>*

- 10) Infant mortality rate
- 11) Under 5 mortality rate
- 12) Maternal mortality rate

### *Natural disasters*

- 13) Whether natural disaster reduced production for more than 80%
- 14) Whether natural disaster reduced production for 50- 80%

### *Land use and agricultural production*

- 15) Amount of cultivate land per capita
- 16) Share of irrigated land
- 17) Grain production per capita

---

<sup>10</sup> Such indicators were proposed by Dr. Wu Guobao who checked the reliability of these data collected by the village doctor.



## *Other*

### 18) Type of projects carried out in the village

Both for the household and village indicators we prepared a very preliminary list of questions that were asked to some households in two different villages and at the village committees respectively.

On average the household interviews lasted 10 minutes. The indicators that are used to predict consumption expenditure in the econometric model should maintain the same definition used in the RPMS and particular attention should also be put in the design of the household questionnaire to follow as much as possible the way in which questions were asked in the RPMS. For instance, in order to get the proportion of cultivated land used for cash crops, it will be necessary to ask how much land is used for each crop grown by the household and then compute the relevant proportion of land used for cash crops (this is how information is collected in the RPMS). Finally, since the econometric results could be reviewed following the suggestions made in the previous section<sup>11</sup>, provided that the information is available in the household survey, it will be important to include in the model any extra variable that we identified in the proposed indicators but that was not among the list of the indicators singled out by the regression analysis (for instance water source and fuel used for cooking).

For what concerns the village level indicators, it is important to note that some of the indicators are already collected by the village statistician (population, land use, etc.) and the village doctor (number of births, number of children deaths and maternal deaths), so it would be relatively straightforward to include information from these sources in a community questionnaire.

## **Next steps**

### **Finalising the list of indicators**

In order to finalise the list of indicators it will be important to further test the results of the regression as suggested in the previous section. Such further analysis is likely to make some changes in the list of indicators. Also it is likely that after further consultations there will be pressures to include more indicators, but it will be important to cross-check that all indicators are really needed. For each indicator it is important to answer the following questions: what is its specific use? Is it addressing the purpose of poverty monitoring? Is such information available from other sources?

Once the final list of indicators is ready it will be necessary to design the questionnaire that collects the information required to compute such indicators. For the variables that are used to predict consumption expenditure and then poverty levels it is important to make sure that the definitions and the way of asking the relevant information follows as much as possible those used in the RPMS. If the indicators will be used to predict poverty it will also be important to test whether the information collected through this new questionnaires are comparable to the variables in the RPMS (one effective way to test this could be to administer the new questionnaire to the 200 households interviewed by the statistical office in Dayao county for the RPMS).

---

<sup>11</sup> Within the indicators identified by the regression analysis, some variables have been omitted from the proposed list because such indicators are likely to be revised once the improvement suggested in the previous section will be carried out (for instance whether the household made substantial home improvements in the last 12 months).

## **Final recommendations**

As mentioned in my previous report the conditions for institutionalising a community based poverty monitoring system are already present in Dayao. Each village has a village committee with people that have specific tasks of data collection (for instance village statistician and village doctor) and such resources could be used to collect such data. However, there are specific issues that need to be carefully investigated to make sure that the proposed indicators can be collected, properly managed and well used for decision making. Although the data collection mechanism and its feasibility will be explored and assessed in later stages of the project I want to point out three general issues that should be given special attention in future work.

1) What is currently lacking is the ability to manage properly the current data. Data is not entered in computers and properly shared and analysed. It would be necessary to investigate what it would be required to make such improvement, and whether the current resources could allow such improvement.

2) I am also of the impression that collecting household level data could be too demanding for the village (the collection of household data could require up to a person month work in each village, and in some cases the presence of many natural villages, some of them not easily accessible, poses specific challenges). A possible alternative could be to collect such data for a sample of households rather than for all households, although the way in which such sample should be selected will have to be carefully assessed.

3) It will be very important to institutionalise an independent mechanism to check the quality of the data.

Finally, I want to stress once more that changes will need to be gradual in order to be effective and sustainable. Important lessons can be learned from the weaknesses of similar attempts to collect data at the village level by the Poverty Reduction Office.

## Appendix A: The econometric model

The following table shows the results of the OLS regression using population weights. The same variables are also highly significant if we correct for the clustering effect.

**Table A.1. Predictors of the logarithm of per capita consumption expenditure**

	Coef.	Std. Err.	T-value
Household size	-0.556	0.029	-19.49
Squared of household size	0.038	0.003	11.19
Whether extended household (3 generations)	0.148	0.016	8.99
% of household members 6 years and under	-0.318	0.052	-6.08
% of household members over 60	-0.275	0.041	-6.76
Whether highest edu. of mem. 15-59 is:			
No education	-0.157	0.028	-5.52
Primary	-0.088	0.014	-6.4
Senior middle or above	0.122	0.025	4.79
Whether household head cannot speak Chinese	-0.156	0.033	-4.79
Whether household has difficulty in receiving medical attention	-0.131	0.017	-7.9
Clinic/hospital is too far to receive medical attention	0.128	0.027	4.81
% of household crop area used for cash crops	0.302	0.045	6.65
% of household labour employed full time in agriculture	-0.111	0.023	-4.9
Whether a concrete house structure	0.150	0.031	4.81
Whether household has telephone	0.154	0.020	7.88
Whether household has colour TV	0.133	0.013	9.93
Whether the household has motorbike	0.229	0.028	8.13
Number of big animals owned by household	0.031	0.008	4.12
Number of pigs owned by the household	0.040	0.004	11.21
Whether one or more household members attend university	0.465	0.047	9.96
Whether household made big home improvements	0.244	0.032	7.62
Whether village is located in hilly area	0.112	0.021	5.37
% of natural villages with electricity	0.219	0.035	6.27
Whether natural disaster reduced production by more than 80%	0.143	0.038	3.73
Whether natural disaster reduced production by 50-80%	0.091	0.022	4.11
Constant	8.222	0.067	123.24
Number of observations		7300	
R-squared		0.464	
Root MSE		0.449	