



PMT-based social registries

Measuring income and poverty using Proxy Means Tests

OVERVIEW

This document will serve as an introduction to proxy means testing (PMT). The concept of proxy means testing is usually understood as using observable characteristics of the household or its members to estimate their incomes or consumption, when other income data (salary slips, tax returns) are unavailable or unreliable.

As the name suggests, the obtained income estimates are in turn used for means-testing, that is to determine whether a household, or an individual, is eligible for support from social safety net programs. If their estimated income falls below a certain threshold, they are deemed eligible for support; if the income is higher than the threshold, the support is withdrawn.

Throughout this guide, we will look at other possible reasons for employing the PMT methodology, the techniques used to create and apply PMT models and look at some of the challenges associated with practical PMT application.



What is the Proxy Means Test?

Proxy Means Test allows us to estimate the income or consumption when precise measurements are unavailable or difficult to obtain. In many situations, we might not be able to tell how much a family earns or spends every month. Even the household members themselves might not be able to tell—they seldom maintain detailed records.

However, we can make an informed guess based on the household characteristics. We know that, all things remaining equal, a family living in a brick-walled house will likely be able to afford more compared to a family living in a house made of clay. The type of wall is our “proxy”, because we can use it to approximate our value of interest—the household income or consumption.

Naturally, using only one proxy value will render our estimations very imprecise—there is still great variability in incomes, even between families living in houses made of clay. But we need not limit ourselves to only one proxy. If we see that, on average, richer families do not only live in brick-walled houses, but also own livestock, such as cattle, we could add “livestock ownership” as another proxy variable. And we need not stop there; we can imagine adding layers of other proxy variables, until we obtain a **PMT model**—a set of variables and weights associated with them that accurately predicts the welfare of different household. In practice, most PMT models use more than a dozen of different variables.

Why do we need to measure income/consumption?

The name **proxy means test** suggests that **means testing** (making sure that the social safety net benefits will only be distributed to households that are poor enough to meet certain income criteria) is its main purpose.

This is true and the use of PMT for **targeting** is something that we will focus on here. It should be noted, though, that the same model can have a few other uses. For example, comparing PMT scores over time and between regions can help in assessing effectiveness of social policies. Since

PMT questionnaires are usually shorter and easier to administer than full-fledged consumption surveys, we can also use them to get a quick snapshot of poverty rates in a given locality and at a given time.

Constructing the PMT model

In the previous example, we have decided that the expected consumption of families with brick-walled houses will be higher compared to the ones with clad-walled dwellings; in other words, brick houses are constructed by more affluent people. While the statement will probably be correct, there are two important parts that we have missed:

- We do not have a proof; so far, our statement has only been based on our intuitions;
- Even if we know that a family with a brick-walled house will be more affluent than the one from a clay dwelling, we do not know by how much; will they consume goods and services worth twice as much? Or three times as much? We could try to come up with a number, but it will still be based merely on our own ideas and intuitions.

A formal PMT model solves those two problems by being based on actual quantifiable data. The data, most frequently, comes from household surveys. Household surveys register both household characteristics (such as the type of dwelling, the type of assets that the household owns, composition of the household, as well as ages, genders and occupations of its members).

Because the surveys tell us about both household characteristics **and** household consumption, they allow us to use statistical methods to find out about relationships between the two. Multiple regression allows us to estimate the relationship between variables describing the household characteristics and the monetary value of household consumption. The following table illustrates the example result of such an analysis¹:

¹ The values are fictitious and do not correspond to the PMT model actually used by any country

<i>variable</i>	<i>coefficient</i>	<i>variable</i>	<i>coefficient</i>
<i>Household owns the house</i>	100	<i>Number of household members in employment</i>	50
<i>One child in the household</i>	40		
<i>Two children in the household</i>	30	<i>Number of persons per room</i>	-20
<i>Three or more children in the household</i>	20		
<i>Household owns cattle</i>	200		
<i>Household owns a bicycle</i>	300		
<i>Household owns a car</i>	800		
<i>Dwelling walls made of brick</i>	100		
<i>Dwelling walls made of tin</i>	0		
<i>Dwelling walls made of clay</i>	-100		
Constant	1000		

In this case, we simply have to multiply the coefficient listed in the table by the value applicable to the household. For example, in a household where two persons are employed and, on average, three persons share a room, we modify the baseline by:

$$2 * 50 + 3 * (-20) = 100 - 60 = 40$$

In the case of this household, we increase their estimated consumption by 40 takas.

Those results allow us to be more precise: let us assume that our household:

- Is composed of two adults and three children;
- Owns its house;
- Owns a bicycle, but not a car;
- Doesn't have any cattle;
- And that the walls of the house are made of clay.

In order to calculate the estimated consumption expenditure of the household, we should: Start with the value listed in the "constant" row. This will be our baseline. In our case, this amounts to 1000 takas (Bangladeshi currency)

- Find all rows that apply to the household situation and add or subtract the corresponding coefficients to or from the "baseline" value.
- Consequently, the estimated value of consumption for the example household will be equal to 1000 (baseline) + 100 (house ownership) + 20 (three or more children) + 300 (ownership of bicycle) – 300 (clay walls) = 1320 takas.

Not all of the variables will refer to simple yes/no questions. It could be the case that an entry in the table will read:

Choosing data source

As we mentioned, the essence of the proxy means test method is using household characteristics to predict household welfare. Therefore, we need a data source that would contain information about both household characteristics (household composition, dwelling qualities, assets) and household welfare (normally measured by consumption volume), and that would allow us to draw correspondence between the two.

This means that we normally need results of a household survey. Household Income and Expenditure Surveys (sometimes called household economic surveys) or Living Standards Measurement Study surveys are examples of surveys that contain data that could be used in developing a PMT model.

Regardless of the exact type of the survey that we want to use, there are two important criteria that our data source has to meet:

- The data should be as recent as possible—as the living conditions of the population change, the relationship between particular household characteristics and its welfare is also prone to change. Ideally, the survey should have been conducted in the last five years.
- The survey results should be representative for the area. If we want to use the PMT scores for

targeting social safety nets beneficiaries nationwide, we should use a survey that is nationally representative.

Choosing the variables

Results of a typical LSMS survey will contain hundreds of variables. Among them, we need to choose the ones that will be used in the PMT model. While doing that, we should take into account two basic criteria:

- “Verifiability” of the variable
- and correlation between the variable and household welfare (consumption levels).

Good PMT variables will be easy to verify by enumerators or social workers. For example, the type of walls in the dwelling can be easily discerned by the enumerator visiting the household—brick is visibly different from wood or tin. This is generally the case with other variables describing housing quality. Although household characteristics (number of members, their ages, education or occupation) could be misrepresented, this does not happen often, in particular when the enumerators or social workers belong to the same local community. Caution should be exercised with including ownership of small assets in the model—these could be removed or concealed.

The relationship between the selected variables can be determined through regression analysis; while comparing results of regressions with different variables included or omitted, the authors of the PMT model can compare R^2 and statistical significance of the coefficients for different models. Once one model has been chosen, the regression coefficients will become the coefficients for the PMT model—they will determine the extent to which each variable affects the household consumption.

Validating the PMT model

After choosing the model, it is important to see how it performs in identifying the poorest households. This is often referred to as ‘**validating**’ the PMT

model. The simplest way of validating the model involves using the data from the same survey that has been described above. Next to the survey, we will also need to know the **poverty line**. In its absence, we can select a **percentage** and assume that the poor comprise the bottom 30 or 40% of households with lowest consumption.

Following that, we conduct the validation by:

- using the survey data, calculate PMT scores for every household in the survey dataset;
- listing the households below the poverty line, according to their consumption (not the PMT score);
- listing the households below the poverty line, according to their PMT scores;
- comparing the two lists.

Since the actual consumption data is more reliable than the PMT data, if a household was included in the consumption-based list, but not in the PMT-based list, we speak of **exclusion error**. The household was poor, but the PMT model has not identified it as such.

On the other hand, if a household was not identified as poor according to its consumption measured by the household survey, but the PMT model has identified it as poor, we speak of **inclusion error**. The household was not poor, but the PMT model considered it poor.

Ideally, we would like to minimize both inclusion and exclusion errors. In practice, however, there is often trade-off between the two, with typical PMT models featuring inherent 30-40% inclusion and exclusion errors.

Different models

It is not required that a single model is used for entire country. Sometimes it might improve accuracy of prediction if separate models for different regions are created. Often a separate model for urban and rural poor is developed, as welfare indicators differ between cities and the countryside (for instance,

livestock ownership, which often accurately predict welfare in the countryside, is irrelevant for the city dwellers).

More advanced PMT models can use multi-staged approach, where the results of one model's prediction are 'plugged in' to another model, in order to minimize inclusion and exclusion errors.

Updating the model

As societies change, PMT models may become outdated. Intuitively, the models can and should be updated when more recent data become available. The household surveys used to produce the models (HEIS, LSMS) are often released in 5- to 10-year cycles, which suggests a similar pattern for PMT model updates.

Bangladesh experience

To finish this note, it may be useful to present the experience of building the PMT model for Bangladesh to see how the principles described above can be applied in practice.

In terms of the data source, the mode has been based on the results of the most recent 2010

Household Expenditure and Income Survey.

The main rationale behind the choice has been

good data quality and the fact that quite details household characteristics variables were included. In addition to that, the survey was used to compute official poverty figures and poverty maps for the country; consequently, its use meant that a consistent definition of poverty could be applied across the board.

The country uses a single model, although the final score is adjusted based on the region. That approach simplifies the model, which is important when it comes to promoting its use across different government agencies. Most of the variables selected for the models can be directly observed by the enumerator visiting the household, minimizing the risk of fraud or misrepresentation.

The variables used by the Bangladesh model and the corresponding scores are included in the following table. Undercoverage, for the cut-off point of 40%, is equal to 36%. While the value could still be considered a sizeable exclusion error, it has been proven through pilots that targeting benefits based on the model is still more accurate compared to the methods previously used for social safety net programs in Bangladesh.