American Economic Review: Papers & Proceedings 2009, 99:2, 87–92
http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.2.87

DATA WATCH: DATA INITIATIVES[†]

Expectations and Perceptions in Developing Countries: Their Measurement and Their Use

By Orazio P. Attanasio*

In recent years, much attention has been devoted to the debate between two groups of applied researchers. The first stresses the need for empirical work to rely on variation in the data that is truly and surely exogenous to estimate a well defined, if narrow, set of parameters. The second is willing to impose structure on data to estimate behavioral parameters. The conduct and design of economic policy rely heavily on the identification of structural behavioral parameters, as they are crucial to understanding how individuals react to incentives in different contexts. However, the estimation of structural models often relies on very strong assumptions. The availability of rich datasets that include credible measurements of variables that are relevant to the decision-making process, such as expectations, perceptions, and beliefs, can shift the weight of this debate, as it allows the estimation of structural models of individual behavior using much weaker assumptions. Moreover, in the case of development, measurement of this type of variables goes straight to the core of issues that are fundamental for the understanding of the causes of the lack of development and the imperfections (in markets, knowledge, and information) that may prevent growth. The availability of hard data on expected returns on certain investments, and how individuals act upon these expectations, gives direct information on whether individuals have reasonable beliefs on returns and, crucially, on whether credit and insurance market imperfections might be playing an important role in determining actual investment behavior.

In this paper, we review recent progress on the measurement of this type of variables in developing countries, and discuss possible future developments.

I. Expectations: Measurement and Use

There is now a small literature on the measurement of subjective expectations. Following a number of early attempts, it has become increasingly clear that, if enough care is devoted to the design of questionnaires, it is possible to elicit high quality information about the probability distribution of future variables that are important for economic welfare and are relevant to determine economic choices. This point was made forcefully by Charles Manski (2004) in his Frisch lecture. Now many examples of subjective expectations of future variables, ranging from income to returns, education, and stock market returns, exist in the literature.

In developing countries the collection of expectations data poses somewhat different challenges but also affords important opportunities. The respondents of surveys in developing countries have often very limited formal education and can be very unfamiliar with the formal concepts of probability (unlike respondents in developed countries who might have been exposed to the concept of probability on a much more regular basis, for instance through weather forecasts). On the other hand, data collection is typically much cheaper in developing countries and, typically, respondents are willing to devote a bit longer to answering surveys than in developed countries.

A paper by Adeline Delavande, Xavier Giné, and David McKenzie (2008) surveys some of the recent contributions to the literature on

[†] *Discussants:* Adeline Delavande, Universidade Nova de Lisboa; Lise Vesterlund, University of Pittsburgh; James Smith, Rand Corporation; Catherine Eckel, University of Texas, Dallas.

* University College London, Gower Street, London WC1E6BT, UK, Institute for Fiscal Studies, and NBER (e-mail: o.attansaio@ucl.ac.uk).

the measurement of subjective expectations in developing countries. Providing evidence from many studies, Delavande et al. (2008) make it very clear that the elicitation of the probability distribution of future variables is not only feasible but is strictly preferred to the elicitation of point expectations and to the use of qualitative scales such as the Likert scale.

In what follows, I will not provide an exhaustive survey, but rather will discuss some issues that are currently still unresolved or are of particular relevance to the collection and use of expectations data in developing countries.

A. *Measurement Tools of Expectations Data*

The fact that survey respondents in developing countries have typically very limited formal schooling makes the collection of subjective expectations data that make use of the concept of probability particularly challenging. The experience of many researchers, however, indicates that such an endeavor is possible if enough care is given to the design of the questionnaires. Moreover, some common protocols that have been proven effective are slowly emerging. In the case of a discrete variable, one typically asks the probability of a given realization. Examples of discrete random variables that are common in the literature are questions about surviving to a certain age and about the probability of unemployment. We come back to the issue of how to ask questions about perceived probabilities below.

In the case of continuous variables, the elicitation of the probability distribution of these variables is obviously harder. Many of the available questions on subjective expectations of continuous variables start with the elicitation of the range of variation of the relevant variable. Respondents are asked to assess what are the "minimum" and "maximum" values a given variable can take at a future date. The wording of these questions should be precise about the appropriate conditioning. These questions are typically reasonably well understood. There is an issue, however, about whether respondents literally interpret the "min" and "max" in these questions. Delavande et al. (2008) provide some interesting evidence showing that, at least in the specific context they analyze, respondents with "the maximum" seem to mean some high percentile.

Having obtained the range of variation for the variables of interest (which can already provide both a measure of location and of variation for the variable of interest), the interviewer typically divides the interval in two or more subintervals and asks questions to assess the probability the respondent attributes to each subinterval. In most cases, two or four subintervals are considered.

The questions about probability are typically the most difficult to ask. Because many of the respondents in developing countries have not been exposed to the concept of probability, it is important to try different versions and different methods of asking the questions. Typically the use of examples (for instance, the probability of rain the next day) is useful in explaining the concept, as is the use of visual aids. Delavande and Hans-Peter Kohler (2008), for instance, use a pile of ten stones to represent probability units. Attanasio, Costas Meghir, and Marcos Vera-Hernández (2005), instead, use a ruler graded from 0 to 100 to which respondents can point to indicate their probability assessments. Experience has shown that no single method works everywhere and that researchers have to be inventive in adapting to the local context and make extensive use of trials. Delavande et al. (2008) mention several examples from different contexts.

An extensive literature exists in psychology and statistics on the best methods to elicit probabilities, although often the focus is on how to elicit probabilities from experts, rather than survey respondents. A large number of issues arise in the elicitation process, for instance, those induced by anchoring, that is, the fact that the framing of a question, such as the mention of specific intervals, affects the way the question is answered by respondents. How these issues translate in the context of developing countries is not clear.

B. *Distributional Assumptions: Converting Measurements into Moments*

While in the case of discrete random variables, the probability measurements give everything that a researcher might want to observe; in the case of a continuous variable, the range of variation and the few probability questions give some point of the cumulative distribution function (CDF) of the relevant future variables. If a researcher plans to use the information elicited from respondents to model behavior, it is likely that she will be interested in specific moments of the probability distribution of the variable of interest, such as

TABLE 1

	Sample average probability that income is		
	Above mid-point	Below mid-point	Sum
Entire sample (N=1,813)	0.4809	0.5931	1.0741
	(0.018)	(0.0106)	(0.0204)
Dropping observations with probability of 0 or 1 (N=1,533)	0.4847	0.5387	1.0234
	(0.018)	(0.0082)	(0.0203)

Note: Standard errors in parentheses.

Source: Attanasio, Meghir, and Vera-Hernandez (2005).

the mean or the variance. Alternatively, within the framework of a structural model, one might want to use the entire distribution. Either way, to use information on subjective expectations, one needs to make assumptions about the nature of the distribution and then use the information on the points of the CDF to characterize them. Several alternative assumptions have been used in the literature, such as log-normality, piecewise linearity, and triangular.

The assumption on the distribution, especially when very few points of the individual CDF are observed, is obviously arbitrary. For this reason, it seems advisable to check the extent to which results are affected by alternative assumptions (see, for instance, Attanasio and Vincenzo di Maro 2008; Attanasio and Katja Kaufman 2008).

C. *Validation of Subjective Expectations Data*

Given the difficulties in eliciting subjective expectations, it is important to validate the data one collects. One might want to use both internal and external validation. As for internal validity, one can check whether the implied moments derived from the subjective expectations variables covary with observed characteristics of the respondents in a way that is consistent with other available information. Attanasio and di Maro (2008), for instance, check how the mean and variation of future household income, derived from data collected within the survey for the evaluation of the Mexican conditional cash transfer programme *Oportunidades,* covary with the education achievement and ethnicity of respondents (more educated and non-indigenous individuals have higher expected income). Similar tests are reported by Attanasio and Kaufman (2008). Attanasio and di Maro (2008) also report that the variability of future income implied by the subjective expectations data covaries with the variability of past income reported by respondents.

In addition to these simple tests, it is relatively easy to build in mechanisms to check the internal validity of the questions. A good example is the one in Attanasio, Meghir, and Vera-Hernandez (2005) (AMV). The survey they study was carried out to evaluate the impact of a workfare programme in urban Colombia. The respondents were asked to state the maximum and minimum expected future income, and the implied range was divided into two intervals. The sample was then split randomly, and half were asked the probability that future income would be between the minimum and the midpoint, while the other half were asked the probability that future income would be between the midpoint and the maximum. One way to validate the expectations data is to test the hypothesis that the sum of the probabilities that income is below and above the midpoint sum up to one. Table 1 reports some of the results in AMV.

The sum of the average probabilities equals 1.07, which is significantly above 1. However, if one drops observations that answer zero or one to the probability question, and those that are clearly inconsistent (given the answers on the minimum and maximum), the sum of probabilities equals 1.02 and is not significantly different from one. Interestingly, at least for this dataset, the fact that the two probabilities sum to one is not driven by most participants replying 0.5 to the relevant question. AMV reports that only about 15 percent of respondents answered 0.5.

Checking external validity of the subjective expectations data is obviously trickier. If data on the realization of income expectations are available for many periods, one can test whether expectations and realizations are systematically

different. However, the data requirements for such an exercise are formidable: one rarely has data on realizations that match the expectations previously held. Even harder is the requirement that such data be available for multiple periods, which is essential if one wants to test the hypothesis of rational expectations in the presence of aggregate shocks. Moreover, while testing the hypothesis of rational expectations is obviously interesting and important, one of the points of having reliable data on subjective expectations is precisely the possibility of doing empirical work *without* having to assume rational expectations.

Analogous considerations arise in other contexts. For instance, Jeffrey Dominitz and Manski (1996) and Attanasio and Kaufman (2008) use data on expected returns to education in Wisconsin and Mexico, respectively. Both studies elicit the probability distribution of future earnings of high school students under different scenarios about their schooling. These data can be used to estimate the expected return to schooling, cutting through the selection issues whose study has generated an entire literature. Validation of these data, by comparison, with actual realization is particularly tricky. Attanasio and Kaufman (2008) do compare the respondents' expected earnings with the actual earnings of 25-year-olds in different datasets. They point out, however, that there are many reasons for subjective expected returns and "realized" returns, defined as the difference between earnings of individuals with different education levels, to differ. The most important is probably selection into education of individuals with different degrees of ability.

D. *Using Subjective Expectations Data*

As the elicitation of subjective expectations data is recently new, not many studies that have looked at these data use them within behavioral models. But as this type of data become more common and accepted, more studies make use of them. Luigi Guiso, Tullio Jappelli, and Daniele Terlizzese (1996) construct measures of income uncertainty from subjective expectations to study portfolio allocations in Italy, while Luigi Pistaferri (2001) uses the same data to identify income shocks and to construct an ingenious test of the permanent income hypothesis (PIH).

On some occasions, the subjective expectations questions refer to a choice variable, such as questions about the probability of retiring at a certain age. In a recent paper, Wilbert van der Klaauw and Kenneth Wolpin (2007) use observations on stated probability of retirement within a structural model of retirement choices.

The applications in developing countries are less numerous, but growing. Delavande et al. (2008) report several examples where moments derived from subjective expectations data are shown to predict and explain actual behavior. Examples range from migration decisions (David McKenzie, John Gibson, and Steven Stillman 2007) to production decisions in Uganda and India (Ruth Vargas Hill 2006; Xavier Giné, Robert Townsend, and James Vickery 2008) to the supply of credit in India's fisheries (Giné and Stefan Klonner 2007) to education choices in Mexico (Attanasio and Kaufman 2008).

Attanasio and di Maro (2008) use the data on income expectations to estimate a model of income dynamics in rural Mexico. The properties of individual income processes are obviously important to understanding consumption, saving, and investment behavior. In particular, the persistence of income shocks has received much attention in the literature. Suppose that the individual income process is given by $y_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + u_{i,t}$ where the shock u is assumed to have zero mean and to be i.i.d. Under rational expectations, subjective expectations would then be $y^e_{i,t} = E[y_{i,t} | y_{i,t-1}] = \alpha_0 + \alpha_1 y_{i,t-1}$.

The first moments of future income derived from subjective expectations data, which we label $y^{ee}_{i,t}$, differ from actual expectations because of measurement error. We can estimate the persistence parameter running the regression

$$y^{ee}_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + v_{i,t}, \tag{1}$$

where v is *not* an income shock but, rather, the measurement error in income expectations. If expectations data are available, equation (1) can be estimated on a single cross section.

There are many reasons why the estimation of equation (1) would yield biased estimates of the persistence parameters, besides violation of the assumption of rational expectations. In particular, if the intercept parameter is individual specific, that is, if there are fixed effects,

the residuals will include the fixed effect and, therefore, lagged income will be correlated with the residuals. A possible solution is to subtract from equation (1) $y_{i,t-1}$ to get

$$y_{i,t}^{ee} - y_{i,t-1} = \alpha_1(y_{i,t-1} - y_{i,t-2}) + v_{i,t} - u_{i,t-1}. \tag{2}$$

Notice now that the residual of equation (2) includes both measurement error and income shocks, so that OLS estimates are likely to be biased. It is, however, possible to employ a GMM or IV strategy, using $(y_{i,t-3} - y_{i,t-4})$ as instruments for $(y_{i,t-1} - y_{i,t-2})$. The problem with this strategy, however, is that we lose the ability to estimate the parameter of interest with a single cross section. Attanasio and Britta Augsburg (2008) propose a solution of this problem by using data on "usual" income to model fixed effects.

II. Perceptions and Beliefs

It has been observed that in many situations poor individuals do not seem to engage in activities with potentially high returns and relatively low costs. A plausible explanation might be the limited information available to poor households. Esther Duflo (2006), for instance, cites a randomized trial in India where the take-up rate of basic vaccination offered at no cost in some rural settings was increased dramatically by the offer of a small in-kind incentive (a kilo of lentils). Poor information about the effect of vaccination could easily explain this type of behavior. But if this type of phenomenon is common, it is important to collect systematic and standardized data on the information and beliefs that people rely upon when making important investment decisions. Child care, nutrition, health care, schooling and education, and agriculture production are all areas in which data on information and beliefs can be very important and which can and should be collected systematically in household surveys.

This agenda does overlap, to an extent, with the measurement of subjective expectations: the formulation of expected returns education requires explicitly expectations about future variables. In other cases, however, the issues are of a different nature and concern specific knowledge of technology parameters.

III. Asymmetric Information

One of the main developments in economics in the last 50 years has been the analysis of environments in which information is distributed asymmetrically. From an empirical point of view, however, the evidence on asymmetries of information and their importance has always been, so far, indirect. The advances in the measurement of subjective expectations discussed in Section I offer the possibility of developing measurement tools that could be used, especially in relatively simple economies, to assess quantitatively the importance of asymmetric information.

Consider, for instance, a village economy where individual incomes are, to an extent, private information. It is well known that the presence of important aggregate shocks, such as the weather, does not necessarily imply that idiosyncratic shocks are unimportant, as aggregate shocks can have different effects on different individuals. What is not fully understood is the extent to which individual shocks are common knowledge or private information.

A well-developed module to collect information on subjective expectations can be profitably used to gather information on the relevance of asymmetric information. In addition to questions about a respondent's own future income, one can think of asking questions about the future income of the respondent's neighbors and fellow villagers, and apply the same questions to the other villagers. A respondent's uncertainty (as measured, say, by the variance) about her own income should be smaller than the other respondents uncertainty about the same variable.

In situations in which income flows are endogenous and depend on privately observed effort, modifications of the methods sketched above are probably necessary.

IV. Conclusions

In this paper, we have argued that while the elicitation of subjective probability distribution of future uncertain variables is not easy, with enough ingenuity it is possible to include questions and modules in standard household surveys that can be used in a variety of situations, in particular, to facilitate the identification of less restrictive behavioral models.

We have also argued that similar considerations apply to other variables that are important for our understanding of economic behavior, particularly so in developing countries. Data on the information, beliefs, and perceptions upon which individuals rely to make important decisions can and should be collected. We need a better understanding of how these beliefs affect behavior and the response to incentives, and, in turn, how they are affected by policies and, more generally, by the economic environment. For this research agenda to grow, it is essential to have enough appropriate measurement tools, which in some cases, e.g., the measurement of the importance of asymmetry of information, have to be developed.

Developing countries constitute an important environment in which this research agenda can and should be promoted. Data collection in these countries is typically much cheaper than in developed economies. Moreover, it is sometimes possible to isolate relatively simple economic environments in which some concepts are relatively straightforward. Last but not least, the sort of problems that can be studied with hard data on expectations, information, and beliefs are particularly salient for developing economies and for the development process.

REFERENCES

Attanasio, Orazio, and Britta Augsburg. 2008. "Subjective Expectations and Investment in Rural India." Unpublished.

Attanasio, Orazio, and Vincenzo di Maro. 2008. "Income Expectations in Rural Mexico: Measurement and Models." Unpublished.

Attanasio, Orazio, and Katja Kaufmann. 2008. "Educational Choices, Subjective Expectations, and Credit Constraints." Unpublished.

Attanasio, Orazio, Costas Meghir, and Marcos Vera-Hernández. 2005. "Elicitation, Validation, and Use of Probability Distributions of Future Income in Developing Countries." Paper Presented at the 2005 Econometric Society Meeting.

Delavande, Adeline, and Hans-Peter Kohler. 2008. "Subjective Expectations in the Context of HIV/AIDS in Malawi." Unpublished.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital: Results from a Randomized Experiment." *Quarterly Journal of Economics,* 123(4): 1329–72.

Dominitz, Jeff, and Charles F. Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources*, 31(1): 1–26.

Duflo, Esther. 2006. "Field Experiments in Development Economics." In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress Volume 2,* ed. R. Blundell, W. K. Newey, and T. Persson, 322–48. Cambridge: Cambridge University Press.

Giné, Xavier, and Stefan Klonner. 2007. "Technology Adoption with Uncertain Profits: The Case of Fibre Boats in South India." Unpublished.

Giné, Xavier, Robert Townsend, and James Vickery. 2008. "Rational Expectations? Evidence from Planting Decisions in Semi-Arid India." Unpublished.

Guiso, Luigi, Tullio Jappelli, and Daniele Terlizzese. 1996. "Income Risk, Borrowing Constraints, and Portfolio Choice." *American Economic Review*, 86(1): 158–72.

Hill, Ruth Vargas. 2006. "Coffee Price Risk in the Market: Exporter, Producer and Trader Data from Uganda." Unpublished.

Delavande, Adeline, Xavier Giné, and David McKenzie. 2008. "Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence." Unpublished.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica*, 72(5): 1329–76.

McKenzie, David, John Gibson, and Steven Stillman. 2007. "A Land of Milk and Honey with Streets Paved with Gold: Do Emigrants Have Over-Optimistic Expectations about Incomes Abroad ?" World Bank Policy Research Working Paper 4141.

Pistaferri, Luigi. 2001. "Superior Information, Income Shocks, and the Permanent Income Hypothesis." *Review of Economics and Statistics*, 83(3): 465–76.

van der Klaauw, Wilbert, and Kenneth I. Wolpin. 2007. "Social Security and the Retirement and Savings Behavior of Low Income Households." Unpublished.

This article has been cited by:

1. R. V. Hill. 2010. Liberalisation and Producer Price Risk: Examining Subjective Expectations in the Ugandan Coffee Market. *Journal of African Economies* **19**:4, 433-458. [CrossRef]