December 29th 2008.
Preliminary: do not cite without permission

Expectations and Perceptions in Developing Countries:

Their Measurement and Their Use

Orazio P. Attanasio

Department of Economics
University College London
Gower Street
London WC1E 6BT – UK
Tel +44 2076795880
Email: o.attanasio@ucl.ac.uk

January 2009

Keywords: Subjective expectations, household surveys in developing countries

ASSA 2009 Meetings: San Francisco, January 3-5, 2009
Session Title: **Data Watch**
Presiding: MARTIN BROWNING, Oxford University
Discussant: ADELINE DELAVANDE, Universidade Nova de Lisboa

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

Orazio P. Attanasio[*]

The use of microeconomic data has become extremely widespread in applied economics. Household and firm level data are now routinely used not only in labour and industrial organization, but also in macroeconomics. The use that is made of the data is extremely varied, ranging from simple comparison of means in the evaluation literature based on Randomized Control Trials, to the matching of some data moments to calibrate the structural parameters of complex models of individual behaviour to the structural estimation of dynamic optimization models. At the same time, partly as consequences of technology advances, many more data sets are available. And much more detailed and high quality data are being collected. From a methodological point of view, important advances have been made in the techniques aimed at the elicitation in surveys of information about a variety of factors that constitute important inputs in the empirical analysis of economic behaviour. A good example, for instance, is the collection of information on household financial and non financial wealth, which was thought to be a very difficult if not impossible variable to measure accurately in a household survey and, instead, is now collected routinely and satisfactorily in many surveys, thanks to the development and standardization of new interview techniques.

In recent years, much attention has been devoted to the debate between applied researchers who stress 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 and those who are willing to impose structure on data to estimate behavioural parameters. The conduct and design of economic policy relies heavily on the identification of structural behavioural parameters, as they are crucial to understand how individuals react to incentives in different context.

[*] University College London, Institute for Fiscal Studies and NBER

However, the estimation of structural models often relies on very strong assumptions. The availability of rich data sets that includes credible measurements of variables that are relevant to the decision making process, such as expectations, perceptions, belief, can shift the weight of this debate, as it allows the estimation of structural models of individual behaviour 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, information) that may prevent growth. The availability of hard data on expected returns on certain investments, and how individuals act upon these expectations, for instance, 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 behaviour.

In this paper, we review recent progress on the measurement of this type of variables in developing countries and discuss possible future developments. In Section I, we discuss the measurement of subjective expectations, probably the area in which more progress has been made. The issues we touch upon in this area are the refinements of measurement tools, the assumptions that need to be made to make use of the measurements that are typically available and the actual use of subjective expectations to model behaviour. In Section II, we discuss a possible development to use subjective expectations questions to measure the extent of asymmetric information. In Section III, we discuss the measurement of perception and beliefs. Section IV concludes.

I. Expectations : measurement and use.

There is now a small literature on the measurement of subjective expectations. Following a number of early attempts is 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 to education, to 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 un-familiar 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 recent paper by Adeline Delavande, Xavier Giné, and David McKenzie (2008) surveys very well some of the recent contributions to the literature on the measurement of subjective expectations in developing countries. Delavande et. (2008) make very clearly, providing evidence from several studies, that the elicitation of the probability distribution of future variables is not only feasible but to be strictly preferred to the elicitation of point expectations and to the measurement of probability assessments via a qualitative scale such as the Likert scale.

The scope of this paper is not to supply an exhaustive survey, but rather to indicate a number of issues that are currently either still unresolved or of particular relevance to the collection and use of expectations data in developing countries.

A. Measurement tools of expectations data.

As mentioned above, 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, in some situations. The experience of many researchers, however, indicates that such an endeavour is possible if enough care is given to the design of the questionnaires. Moreover, some common protocols that have been proven to be 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 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 and do not require much time to obtain. There is an issue, however, discussed in Delavande et al. (2008) about whether respondents literally interpret the 'min' and 'max' in these questions as such or whether they provide some interesting evidence showing that, at least for the specific context reported, respondents seem to answer some high percentile for the maximum. 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 sub-interval and asks questions to

assess the probability the respondent attributes to each sub-interval. In most cases, two or four sub-intervals are considered.

The questions about probability are typically the most difficult to ask. Many of the respondents in developing countries have not been exposed to the concept of probability and might find it difficult to answer consistently. For these questions, it is important to pilot different versions and different methods of asking the questions. Typically the use of examples (for instance the probability of rain the next day) are useful in explaining the concept. In addition, the use of visual aids can also be very useful. Delavande and Hans-Peter Kohler (2007), for instance, use a pile of ten stones to represent probability units. Orazio Attanasio, Costas Meghir and Marcos Vera-Hernandez (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 methods works everywhere and that researchers have to be inventive in adapting different methods to the local context and make extensive use of pilots before collecting this type of data. Delavande et al. (2008), mention several different ways in which probability questions have been asked in different contexts.

An extensive literature exists in psychology and statistics on what are 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. 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 CDF of the relevant future variables. If a researcher plans to use the information elicited from respondents to model

behaviour, it is likely that she will be interested in specific moments of the probability distribution of the variable of interest, such as 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 the 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. This is the strategy typically followed in the literature. Several alternatives have been used in the literature, such as log-normality, piecewise uniform, 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. For instance, in cases where questions on the probability mass relative to two subintervals, Attanasio and Vincenzo di Maro (2008) and Attanasio and Katja Kaufman (2008), use two alternative assumptions: a stepwise uniform and a triangular distribution for future income in rural Mexico and future earnings of high school students. In both cases the alternative assumptions make little difference for the first moments but, obviously, make a large difference for the second moments.

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 co-vary 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*, co-vary 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) who study the expectations of future earnings (under different scenarios) held by Mexican high school students. Attanasio and di Maro (2008) also report that the variability of future income implied by the subjective expectations data co-varies 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 collected 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 was asked the probability that future income would be between the minimum and the mid-point, while the other half was asked whether the expected income was between the mid-point and the maximum. The test that the sum of the average for the two probabilities elicited in the two samples equals one constitutes a validation test of the subjective expectations. Table 1 reports some of the results in AMV.

Table 1

	Avg.prob. income is above mid-point	Avg.prob. income is below mid-point	Sum
Entire sample (N=1813)	0.4809	0.5931	1.0741
	(0.018)	(0.0106)	(0.0204)
Dropping observations with prob. of 0 or 1. (N=1533)	0.4847	0.5387	1.0234
	(0.018)	(0.0082)	(0.0203)

Source: Attanasio, Meghir and Vera-Hernandez (2005). Standard errors in parenthesis.

The sum of the average probabilities equals 1.07, which is significantly above 1. However, if one drops observations that answer 0 or 1 to the probability question and 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 the data set, the fact that the two probabilities sum up to one is not driven by most participants replying 0.5 to the relevant question. AMV reports that about 15% of respondents answer 0.5 and that

if one excludes them from the computations in Table 1, the results of the second raw do not change much.

Checking external validity of the subjective expectations data is obviously trickier. If one has data on the realizations of the income expectations one can check, especially, if one has data for many time periods, whether the expectations held by the individuals are actually rational. However, the data requirements for such an exercise are formidable: one rarely needs data on realizations that match the expectations previously held. Even harder is the requirement that such data is available for multiple periods, which is essential if one wants to test the hypothesis of rational expectations while allowing for the presence of aggregate shocks. Moreover, while testing the hypothesis of rational expectations, which is commonly used in much empirical research, 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 about tests of external validity of expectations data based on the comparison between expected values derived from subjective expectations and 'realizations' in other contexts. For instance, Jeffrey Dominitz and Manski (1996) and Attanasio and Kaufman (2008) use data on expected return to education in Wisconsin and Mexico respectively. Both studies elicit the probability distribution of future earnings (at age 25 in the case of Attanasio and Kaufman and at 30 and 40 in the case of Dominitz and Manski) of high school students under different scenarios about their schooling. These data can therefore be used to estimate the expected return to schooling cutting through the selection issues whose solution has generated an entire literature. But then of course validation of these data by comparison with actual realization is particularly tricky. Attanasio and Kaufman (2007) do compare the respondents expected earnings with the actual earnings of 25 years old in different data sets. However, they point out that even ignoring cohort

effects, there are many other reasons for subjective expected returns to education and observed earning differences between individuals with different schooling to differ, first and foremost, the selection into education.

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 behavioural models. But as this type of data become more common and accepter, 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 construct an ingenious test of the Permanent Income Hypothesis.

In 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, along with the observations referring to actual 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 behaviour. Examples range from migration decisions (David McKenzie, John Gibson and Steven Stillman, 2007) to production decisions in Uganda and India (Ruth Vargas Hill, 2006 and 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 understand consumption, saving and investment behaviour. In particular, the persistence of income shocks has received a lot of attention in the literature. Suppose that the individual income process is given by the following expression:

(1) $$y_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + u_{i,t}$$

where the shock *u* is assumed to have zero mean and to be i.i.d.. Under rational expectations, subjective expectations would then be: $y^{e}_{i,t} = E[y_{i,t} \mid y_{i,t-1}] = \alpha_0 + \alpha_1 y_{i,t-1}$

If we denote the first moments of future income derived from subjective expectations data with $y^{ee}_{i,t}$, they will differ from actual expectations because of measurement error. In such a situation, we can estimate the persistence parameter running the regression:

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

where *v* is *not* an income shock but the measurement error in income expectations. An advantage of equation (2) is that it can be estimated on a single cross section, collected at *t-1*, which includes data on elicited expectations $y^{ee}_{i,t}$.

Of course there are many reasons why the estimation of equation (2) would yield biased estimates of the persistence parameters, beside violation of the assumption of rational expectations. In particular, if the intercept parameter is individual specific, that is there are fixed effects, then the residuals will include the fixed effect and, therefore, lagged income will be correlated with the residuals. A possible solution, then is to subtract from equation (2) an expression for $y_{i,t-1}$ to get:

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

Notice that now the residual of equation (3) includes both measurement error and income shocks, so that OLS estimates are still likely to be biased. It is however possible now, to use a GMM or IV strategy, using lagged changes in income as instruments for the change between *t-2* and *t-1*. The problem with this strategy, however, is that we loose the ability to estimate

the parameter of interest with a single cross section. Attanasio and Britta Augsburg (2008) propose a solution of this problem using data on income and elicited subjective expectations from rural India, by using, in addition to current and expected income, data on 'usual' income to model fixed effects.

II. Perceptions and beliefs.

An interesting use of expectations data is the one recently presented by Robert Jensen (2006). He runs a randomized trial in the Dominican Republic where he provides a sample of poor students with information on returns to education. It turns out that the subjects in the experiment held expectations about returns lower, on average, than returns observed in actual data. The striking result is that the subjects that received the information from the researcher changed their behaviour, in that they were more likely to enrol in school.

While the study has obvious ethical implications, as it is not clear what the 'return to education' for a poor Dominican student is, it brings to the attention an issue, which is probably of first order importance in many developing countries. In many situations in which poor individuals do not seem to engage activities with potentially high returns and relatively low costs, a plausible explanation is the limited information available to poor households. Esther Duflo (2005), 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 incentive in-kind (a kilo of lentils). Poor information about the effect of vaccination could easily explain this type of behaviour. But if this type of phenomena are 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, agriculture productions are all areas in which data on information and beliefs can be very important and that can and should be collected systematically within 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 development in economics in the last fifty years is the analysis of environments in which information is distributed asymmetrically. There have been fundamental theoretical advances that have changed profoundly our understanding. 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 that we have 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 in determining a number of outcomes.

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, instead, 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. One could think of asking a respondent questions about her own future income but also similar questions about the future income of her neighbours and fellow villagers. One can then symmetrically apply the same questions to the other villagers. Respondent A's uncertainty (as measured for

instance by the variance) about her own income should be smaller than the variance of respondent B's uncertainty about A's income.

For situations in which the problem is likely to be pure information about an exogenous income flow, the methods discussed above can be applied directly. In situations in which, instead, the income flows are endogenous and depend on privately observed effort, modifications of the relevant questions are probably necessary.

IV. Conclusions.

In this paper, we have discussed some recent developments in the measurement of subjective expectations and the applications of these methods in developing countries. The main message there is that while the elicitation of subjective probability distribution of future uncertain variables is not easy, with enough ingenuity and care it is possible to include questions and modules in standard household surveys that can be used in a variety of situations and, in particular, to facilitate the identification of less restrictive and more credible behavioural models.

We have also argued that similar considerations apply to other quantities that are surely important for our understanding of economic behaviour and particularly so in developing countries. These quantities include data on the information, beliefs and perceptions upon which individuals rely to make important decisions. We need a better understanding of how these beliefs affect behaviour, the response to incentives and, also, how they are in turn 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, such as 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 pushed. 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. And, last but not least, the sort of problems that can be studied with hard data on expectations, information, 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", In progress.

Attanasio, Orazio and Vincenzo di Maro (2008): "Income expactations in rural mexico. Measurement and models", Mimeo.

Attanasio, Orazio and Katja Kaufmann (2008): "Educational choices, subjective expectations, and credit constraints", Mimeo.

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

Delavande, Adeline and Hans-Peter Kohler (2008) "Subjective Expectations in the Context of HIV/AIDS in Malawi", Working Paper, University of Pennsylvania.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2008) "Returns to capital: Results from a randomized experiment", *Quarterly Journal of Economics*, 123(4): 1329-1372.

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

Duflo, Esther, (2006): "Field Experiments in Development Economics", in Richard Blundell, Whitney K. Newey, Torsten Persson (eds): *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress of the Econometric Society*, Cambridge University Press

Giné, Xavier and Stefan Klonner (2007) "Technology Adoption with Uncertain Profits: The Case of Fibre Boats in South India, Mimeo World Bank.

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

Guiso, Luigi, Tullio Jappelli and Daniele Terlizzese (1996): " Income Risk, Borrowing Constraints, and Portfolio Choice" *The American Economic Review*, Vol. 86 (1), pp. 158-172.

Hill, Ruth Vargas (2006) “Coffee Price Risk in the Market: Exporter, Producer and Trader Data from Uganda”, Mimeo. IFPRI.

Jensen, Robert (2006) **“**The Perceived Returns to Education and the Demand for Schooling”, Mimeo Brown University.

Delavande Adeline, Xavier Giné and David McKenzie,(2008): “Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence” Mimeo

Mahajan, Aprajit, Alessandro Tarozzi, Joanne Yoong, and Brian Blackburn (2008) “Bednets, Information, and Malaria in Orissa”, Mimeo. Stanford University.

Manski, Charles (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 No. 4141*.

Pistaferri, Luigi (2001): “Superior Information, Income Shocks, and the Permanent Income Hypothesis” *The Review of Economics and Statistics*, Vol. 83 (3), pp. 465-476

van der Klaauw, Wilbert and Kenneth I. Wolpin (2007): “Social Security and the Retirement and Savings Behavior of Low Income Households”, Mimeo.